



**THIS PAGE BLANK (USPTO)**

[illegible]

**THIS PAGE BLANK (COPY)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 21, 2002, 22:11:54 : Search time 3431 Seconds

(without alignments)  
4263.378 Million cell updates/sec

Title: US-09-809-638-2

Perfect score: 3720  
Sequence: 1 MTSIMREILLESLLGCYSWS.....EGHNYENNHHFHMTPKFL 699

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+g2n\_model -DEV=xlh  
-O=/cgn2.1/USPRO.spool/US09809638/runat\_21102002\_153504\_1831/app\_query.fasta.1.839  
-DB=GenEmbl -OPT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=humad4.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09809638 -CGN\_1\_1=231 -runat\_21102002\_153504\_1831 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -Delop=6 -Delext=7

Database :  
GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3713	99.8	2486	9 AK025164	AK025164 Homo sapi
2	3125	84.0	2412	10 BC016523	BC016523 Mus muscu
3	760	20.4	31613	8 SCCHRIIT	X59720 S.cerevisia
4	639	17.2	34034	8 SPAC688	AL355632 S.pombe c
5	625.5	16.8	1234	8 DB9130	DB9130 Schizosach
6	495	13.3	995	11 CNS06JWV	AL402149 T3 end of
7	471	12.7	184206	9 AC020593	AC020593 Homo sapi
8	287.5	7.7	222885	2 AL158816	AL158816 Homo sapi
9	268.5	7.2	222885	2 AL158816	AL158816 Homo sapi
10	264.5	7.1	18342	8 SPAC589	AL512496 S.pombe c
11	182.5	4.9	165893	2 AC106576	AC106576 Rattus no
12	182	4.9	305	6 AX337478	AX337478 Sequence
13	152.5	4.1	849	11 CNS06EJ2	AL395172 T3 end of
14	144.5	3.9	225532	2 AC096364	AC096364 Rattus no
15	141.5	3.8	31717	1 SCJ21	AL109747 Streptomy
16	139	3.7	1348	9 HMTPTPSPA	K03223 Human tpi (
17	138	3.7	17013	1 EAMASL	X7921 E.amylovira
18	134	3.6	9720	1 PBDNG	Z73914 Pseudomonas
19	133	3.6	274050	1 AL627276	AL627276 Salimone1
20	132	3.5	29593	8 AF402142	AF402142 Hyalograph
21	131.5	3.5	7330	1 AB026907	AB026907 Ochrobact
22	130.5	3.5	240050	1 AL627267	AL627267 Salimone1
23	130	3.5	193050	1 AJ414149	AJ414149 Yersinia
24	129	3.5	96532	9 CNS07EG6	AL590079 Human chr
25	128	3.4	879	6 AB5835	AB5835 Sequence 49
26	128	3.4	879	6 AR155328	AR155328 Sequence
27	128	3.4	879	6 E65853	E65853 Genome DNA
28	128	3.4	11299	1 AE005976	AE005976 Caulobact
29	128	3.4	20396	1 AE008827	AE008827 Salimone1
30	128	3.4	133863	9 AL157709	AL157709 Human DNA
31	127.5	3.4	206045	2 AL669626	AL669626 Mus muscu
32	127	3.4	4715	1 AF039919	AF039919 Brevibact
33	126	3.4	1340	9 S73482	S73482 Uterine wat
34	126	3.4	26720	1 AE008724	AE008724 Salimone1
35	125.5	3.4	101648	1 AF339846	AF339846 Microscll
36	125	3.4	1340	6 AR002457	AR002457 Sequence
37	125	3.4	1340	6 AR028509	AR028509 Sequence
38	125	3.4	1340	6 AX331552	AX331552 Sequence
39	125	3.4	1340	6 AX331945	AX331945 Sequence
40	125	3.4	1340	6 HUMCHP28A	HUMCHP28A
41	125	3.4	1657	9 H5U41517	H5U41517 Human chann
42	125	3.4	2764	9 BC022486	BC022486 Homo sapi
43	125	3.4	30132	1 SC1068	AL158057 Streptomy
44	124.5	3.3	1647	6 AX120510	AX120510 Sequence
45	124.5	3.3	1770	6 AX066343	AX066343 Sequence

#### ALIGNMENTS

RESULT 1  
AK025164  
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens CDNA: FLJ21511 fls, clone COL05748.  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS Oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens colon CDNA to mRNA, clone\_11b.COL clone:COL05748.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
CDS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (sites)  
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project  
Unpublished (2000)  
2 (bases 1 to 2486)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.  
Direct Submission  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by science and Technology Agency).  
Location/Qualifiers  
1..2486  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Col05748"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/note="Cloning vector pME18SFL3"  
183..2282  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15080.1"  
/db\_xref="GI:10437626"  
Translation="MTSLMKREILLESILGCVSWLYHDLGMYYFPLQTLLELGLGFSIAFLSPILFLITPEFKLVNKKMMLTLRLITIGSISFQAPNAKLLMVALGVSSSLIVQAVTWMSGSHLQRYLRIMGFLIGIVLRLIWTSLNPIINSYMSKMLITLTSALATLRDICTGDCSKPREKRTGATGMAARPNLLAAGSLVFLTHVFGVSLVSRNAVSGHPHPDPNPGCAVLICLASGLMSCLMFRTGLIMWTGTASAGLVLHTMAAVSVCFAITFTASMPQTLGLINSNGPKWTMTAMLEYLEFCAMCTAFKTPGVGVARERSDVLTGTMMLIGLNMLEGPKNKDLLOTNNSKSLVFRSEKYMKLFLMLIVGVLGLGLRKAVERKLGKVAPTKEVSAIIMPFRGYNEGWSLERSAHLNETGADFTLIESDASKPYMGNLUTMGLKGLGFYDGFGSTYTHMGIMALSRYPIKSEHHLISPEDEIATILITVMSKIDVFTVHNGHNDODRLQIALVSKLKSSNOVFLICYITSAKPSRDYLTQITSEGNKVDIDSTDHDRCEYIKRGLIRLVARISHAELSDSEIOMAKFRILPDPTNRDMOKVYIDHREVSERKIHFNPRGSTKEGHNYENNHNHMTPKYFL"  
COUNT 692 a 516 c 606 g 672 t  
Alignment Scores:  
Pred. No.: 1.36e-294 Length: 2486  
Score: 3713.00 Matches: 698  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.86% Mismatches: 0  
Query Match: 99.81% Indels: 0  
Gaps: 0  
US-09-809-638-2 (1-699) x AK025164 (1-2486)  
QY 1 MetThrSerLeuTrpArgGluIleLeuLeuGluSerLeuLeuGluCysValSerTrpSer 20  
DB 183 ATGACCTCGCTGGAGACAAATCTCTTGGAGTCGCTCTGGAGATGTTCTTGGTCT 242  
QY 21 LeuTyrlHisAspLeuGluPrometIleTyrrPheProLeuGluThrLeuGluLeuThr 40  
DB 243 CTCACACATGACCTGGAGACGATGATCTATCTTCTTCCAAACACTACAGCCACT 302  
QY 41 GlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePheLeuThrIleThrProPhe 60

DB 303 GGGCTTGAAGGTTTTAGTATACCAATTTCTTCCAAATATTCCTTAACAATTAATCTCTTTC 362  
QY 61 TrpLysLeuValAsnLysLysTrpMetLeuThrLeuLeuArgIleIleThrIleGlySer 80  
DB 363 TGGAAATTCGTTAACAGAGAGGATGCTAACCCCTGCTGAGATATATCACTTTGGCAGC 422  
QY 81 IleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyVal 100  
DB 423 ATAGCCTCTCCAGGCTCCCAATATCCCAAACTTCACATGATGTTCTTGGCTGGGGTG 482  
QY 101 SerSerSerLeuIleValGlnAlaValAlaThrTrpTrpSerGlySerHisLeuGlnArgTrp 120  
DB 483 TCTTCTCTCATATATGTCACACCTGTGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 542  
QY 121 LeuArgIleTrpGlyPheIleLeuGluGlnIleValLeuValLeuArgIleTrpTrp 140  
DB 543 CTCACAAATTTGGGATTCATTTTACAGACAGATGTTCTTGTGTGTGTGTGTGTGTGTGT 602  
QY 141 ThrSerLeuAsnProIleTrpSerTrpGlnMetSerAsnLysValIleLeuThrLeuSer 160  
DB 603 ACTTCACATAACCAATCTGAGTATCAGATGTCACCAAGATGATCTGACATTAAGT 1662  
QY 161 AlaIleAlaThrLeuAsnArgIleGlyThrAspGlyAspGlySerTrpProGlnLys 180  
DB 663 GCCATGACCACTGATGCTGATTTGGCACAGATGTCAGTGAACCTGAAGAAAG 722  
QY 181 LysThrGlyGlyValAlaAlaThrGlyMetAlaSerArgProAsnTrpLeuAlaGlyAla 200  
DB 723 AAGACTGGAGAGTACCCAGCGGAGTGGCTCTTAACCCAACTGGCTCTGGCAGGGCT 782  
QY 201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGlyValSerLeuValSer 220  
DB 783 GCTTTGGTACCTGTGTGTCTCCACACCACCTGGTTTTGGAGAGTCTCTTGTGTTC 842  
QY 221 ArgTrpAlaValSerGlyHisProHisProGlyProAsnProAsnProPheGlyGlyAla 240  
DB 843 AGATGGCAGTAGTAGGCGATCCACATCCAGGCGCAGATCCATTAACCCATTTGGAGTGTCA 902  
QY 241 ValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCysLeuTrpPheArgGlyThr 260  
DB 903 GTACGCTGTGCTTGGCAAGTGGATGATGCTTCCATCTTGTGTGTGTGTGTGTGTGTGT 962  
QY 261 GlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuTyrlHisThr 280  
DB 963 GGTTCGATCTGGTGGGTAAACAGAACACCTCAGCTGGGGCTCTTACTCGACACAA 1022  
QY 281 TrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSerMetTrpProAla 300  
DB 1023 TGGCAGCTGCTGT 1082  
QY 301 ThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMet 320  
DB 1083 ACACCTGGACACCTTATTAACACAGGACAAACCTGGGAAACCATCACTTCCATG 1142  
QY 321 IlePheTyrlLeuLeuGluIlePhePheCysAlaTrpCysThrAlaPheLysPheValPro 340  
DB 1143 ATATTTCATCTTCTGAATAATATTTTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1202  
QY 341 GlyGlyValTyrlAlaArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIle 360  
DB 1203 GGAGGTGTCTAGCTAGAGAAAGATCAATGCTTTTGGGGCAATATGTTAATATTC 1262  
QY 361 GlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuGluThrLysAsn 380  
DB 1263 GGGCGAATATGCTATTTGGTCTTAAGAAAAACCTGTGATTTGCTCTTCAAAACAAAAAG 1322  
QY 381 SerSerLysValLeuPheArgLysSerGlyLysTyrlMetLysLeuPheLeuTrpLeuLeu 400  
DB 1323 AGTTCTAAAGTCCTTTTCAGAAAGAGTGAAAAATCATGAAGAACTTTTCTGTGCTGCTT 1382  
QY 401 ValGlyValGlyLeuLeuGlyLysLeuArgHisLysAlaTyrlGlyArgLysLeuGly 420  
DB 1383 GTTGGTGTGGATGTTGTGGATTAAGACTACGGCATTAAGCCTATAGAGAAACTGGGC 1442

QY 421 LysValAlaProThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAsp 440  
 |||||  
 DB 1443 AAGTGGACCAACCAAGAGGTCTCTGTCGATCGTGGCTTTCAGGTTGGATATGAC 1502  
 QY 441 AsnGluGlyTrpSerSerLeuGluValArgSerAlaHisLeuLeuAsnGluThrGlyAlaAsp 460  
 |||||  
 DB 1503 AATGAGAGGTGGTCTAGTCTGAGAAAGATCAGCTCAGCTCAATGAAACAGGTGCAGAT 1562  
 QY 461 PheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThr 480  
 |||||  
 DB 1563 TTCATAAACAATTTGGAGAGATGATCTCTAAGCCCTATATGGGACACATGACTTAAC 1622  
 QY 481 MetTrpLeuGlyGluValLeuGluGlyPheThrTrpAspPheGlyProSerThrArgTyrHis 500  
 |||||  
 DB 1623 ATGTGGCTAGGGGAGAAAGTGGGTTCTATACAGACTTTGGTCCAAAGCAAGATATCAC 1682  
 QY 501 ThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeuLeu 520  
 |||||  
 DB 1683 ACTTGGGGATATATGGCTTTGTCAGAGATACCAATTTGGAATCTGAGCATCAGCTTCTT 1742  
 QY 521 ProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnIleSerGlyLys 540  
 |||||  
 DB 1743 CCGTCACAGAGGGCGAGATCGCACACGACATTCACATTGACCGTTAACATTTCCGGGCAAG 1802  
 QY 541 LeuValAspPheValIleThrHisPheGlyAsnHisGluAspAspLeuArgLysLeu 560  
 |||||  
 DB 1803 CTGGGGATTTTGTGTGACACACTTGGGAACCAAGAGATGACCTCGACAGGAACG 1862  
 QY 561 GluAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGluValIlePheLeuGly 580  
 |||||  
 DB 1863 CAGGCTATATGCTGTTCAAACTACTGAAAGTACTCTAATCAAGTAGTATTTCTGGGA 1922  
 QY 581 TyrIleThrSerAlaProGlySerArgAspTyrLeuGluLeuThrGluHisGlyAsnVal 600  
 |||||  
 DB 1923 TATATCACTTCAGCACCTGGCTCCAGAGATATATCTACAGCTCAGCAACATGCAATGTG 1982  
 QY 601 LysAspIleAspSerThrAspHisAspArgTyrPyrGlyGluTyrIleMetTyrArgGlyLeu 620  
 |||||  
 DB 1983 AAGGATATCGACAGCAGCATGATCATGACAGATGGTGTGATATGATATGATGAGGGCTG 2042  
 QY 621 IleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMet 640  
 |||||  
 DB 2043 ATCAGGTGGGTTATGCAAGATCTCCATGCTGAAGTGAAGTTCAGAAATTCAGATG 2102  
 QY 641 AlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGluLysValIle 660  
 |||||  
 DB 2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATATATAGAGACCAACAGAAAGTGTCAATA 2162  
 QY 661 AspHisArgGluValSerGluLysIleHisPheAsnProArgPheGlySerTyrLysGlu 680  
 |||||  
 DB 2163 GACCAAGAGAAAGTCTGAGAAATCTATTTAATCCAGATTTGGATCTTACAAAGAA 2222  
 QY 681 GlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPheLeu 699  
 |||||  
 DB 2223 GGACACAAATTATGAAACAACACATATTTTCATATGAAATACCCCAAAATCTTTTA 2279  
 RESULT 2  
 LOCUS BC016523 2412 bp mRNA linear ROD 05-NOV-2001  
 DEFINITION Mus musculus, similar to hypothetical protein FLJ21511, clone  
 MGC:27925 IMAGE:3584006, mRNA, complete cds.  
 ACCESSION BC016523  
 VERSION BC016523.1 GI:16741399  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2412)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-2001) National Institutes of Health, Mammalian

REMARK  
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [villalobosbcm.tmc.edu](mailto:villalobosbcm.tmc.edu)  
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Hollaway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 35 Row: X Column: 5  
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES  
source

## CDS

location/Qualifiers  
 1. 2412  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="MGC:27925 IMAGE:3584006"  
 /tissue-type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."  
 /clone\_lib="NCI CGAP\_Mam1"  
 /lab\_host="DH10B"  
 /note="Vector: pcMV-SPORT6"  
 113..2212  
 /codon\_start=1  
 /product="Similar to hypothetical protein FLJ21511"  
 /protein\_id="AAH16523.1"  
 /db\_xref="GI:16741400"  
 /translation="MPGLMRALVLETLGVSMSLYGLSPMYPPPLQTELTGLEF  
 FCVAFLEPILTLTIPLMKLVKKWTLISLRIVTVGSIASFEPNAKLRLVLAQVSS  
 SLIVQVYVMWSGLQRLKILWGFILGHVLVLVIVTSINPIMSYOMSNRVLTS  
 AVAILDRIGTGDYDRNPBGKPREVAGRTLSMLPLGAAFGSLFLTHRIFGEVSI  
 YSRVAVSGHPHGPDPGPGAVILGFSGLIMSSMLHDAGLMMTGAASMGILL  
 YLRTMAAVSGCVLAFTGSKMPQYLGHVNSGNSGEAATGAILVLYQTFCAKCT  
 AKFTVPGGVARERSDVLGTTWVILGSLMEGRKNIDELQRRNSPKTLRCSERY  
 MKLIMLFVGVGLGLGRHRTYRQLRGAPVNSAIIWPFRRGVDNEGMPLIERS  
 AOLIKRGADFTILLESAPKPYIGNDITWMLGEKLGFTDFPSTRDHTWGLMVS  
 RPLIVRSFHHLLPDEGEIAPATIMTVNSRLVDPEVYTHRGNEHDLDRLOIAVS  
 KLKNGSNQVTELCYITSEPSRDYIOLTKRGNKIDDSFGDWCEYIWRGLIRLG  
 YARISHAELSDSELOMAKFRILPDPANTRDNQKXVYIDHRGVPKNIHNPREGSKTECH  
 NYENTHHEHMTPEYFV"  
 BASE COUNT 628 a 561 c 616 g 607 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.9e-246 Length: 2412  
 Score: 3125.00 Matches: 579  
 Percent Similarity: 88.70% Conservative: 41  
 Best Local Similarity: 82.83% Mismatches: 79  
 Query Match: 84.01% Indels: 0  
 DB: 10 Gaps: 0

US-09-809-638-2 (1-699) x BC016523 (1-2412)  
 QY 1 MetThrSerLeuTrpArgGluIleLeuLeuGluSerLeuLeuGlyCysValSerTrpSer 20  
 |||  
 DB 113 ATGCCAGGCGCTGTGAGAGCAATTCCTCGAGACTTGTGGCTAGCTCCCTGGTCT 172  
 |||||  
 QY 21 LeuTyrHisAspLeuGlyProMetIleTyrTyrPheProLeuGluThrLeuGluLeuThr 40  
 |||||  
 DB 173 CTCTACCAAGCAGCATGAGCCCATGATCTATTTTCCCTGCAAAAGCTGAGCTCACA 232  
 |||||

QY 41 GlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePheLeuThrIleThrProPhe 60  
 DB 233 GGCTGGGAATTCCTTGTAGCTTCTCTCCGATACCTCTTAAACAATGCCCACTC 292  
 QY 61 TrpLysLeuValAsnLysTrpMetLeuThrLeuLeuArgIleIleThrIleGlySer 80  
 DB 293 TGGAACTGCTCAATAGAAAGTGGACACTCTCTCTGAGATGCTCACTAGGACAC 352  
 QY 81 IleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyVal 100  
 DB 353 AATAGCTCTTGGAGCTCCCAATGCAAGCTTCGACTGATGCTCTCCCTTGGTGG 412  
 QY 101 SerSerSerLeuIleValAlaAlaValThrTrpTrpSerGlySerHisLeuGlnArgTyr 120  
 DB 413 TCTTCTTCTTGTATGCTGCAAACTGCTGCTGCTGCTGCAAGATGGATGCAAAAGTAC 472  
 QY 121 LeuArgIleTrpGlyPheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyr 140  
 DB 473 CTCAAATCTGGGATTCATCTGGGACAGCTCTCTCTCTCTCTCCCAATATGCTAC 532  
 QY 141 ThrSerLeuAsnProIleTrpSerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160  
 DB 533 ACATCTTTAAACCAATTTGGAGCTATCAGATGTCACAGAGATGATTTCTGACATTGAGT 592  
 QY 161 AlaIleAlaThrLeuAsnAspArgIleGlyThrAspGlyAspCysSerLysProGluGlnLys 180  
 DB 593 GCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652  
 QY 181 LysThrGlyGlnValAlaThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAla 200  
 DB 653 AAGCCCGGTAGAGGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 712  
 QY 201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGlnValSerLeuValSer 220  
 DB 713 GCCTTGGGAGGCT 772  
 QY 221 ArgTrpAlaValSerGlyHisProHisProGlyProAspProAsnProPheGlyAla 240  
 DB 773 ACATGGGCGAGTACGTGATCCGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 832  
 QY 241 ValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCysLeuThrPheArgGlyThr 260  
 DB 833 GTCTACTGCTGCTCTCAAGTGGAGTACGTATGCGGTTCAATCAATGAGGCTTATGATGCT 892  
 QY 261 GlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuThrLeuHisThr 280  
 DB 893 GGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 952  
 QY 281 TrpAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSerMetTrpProGln 300  
 DB 953 TGGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1012  
 QY 301 ThrLeuGlyHisLeuLeuAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMet 320  
 DB 1013 GTACTTGGCCACTGTGTAGACTAGGGAACCTCCGGAAGGCCATGGCCACTGGCCATG 1072  
 QY 321 IlePheTyrLeuLeuGlnIlePhePheCysAlaTrpCysThrAlaPheLysPheValPro 340  
 DB 1073 ATCCTTATGCTTTTACAAACATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1132  
 QY 341 GlyGlyValTyrAlaArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIleIle 360  
 DB 1133 GGAGGTGTCTACGCTAGAAAGATCTGATGCTTTTGGGAGCCATATATGTAATATTAT 1192  
 QY 361 GlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuGlnIleThrLysAsn 380  
 DB 1193 GGATTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1252  
 QY 381 SerSerLysValLeuPheArgLysSerGlnLysTyrMetLysLeuPheLeuLeuLeu 400  
 DB 1253 ACTCTAAAGACCTTTTGGAGTACGAGTGAAGAAATACATCAAACTCATTTCTGGCGCTTT 1312  
 QY 401 ValGlyValGlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeuGly 420

DB 1313 GTTGGTGGCTGCTACTGGAATTAGGGCTACGGCACAGAACTACGAGGAGGATTAGGC 1372  
 QY 421 LysValAlaProThrLysLysValSerAlaAlaIleTrpProPheArgPheGlyTyrAsp 440  
 DB 1373 CGAGGGGACACAGCGACACTGGTCTGGCTGCATCTGGCTTTTAAAGTTTGATATGAC 1432  
 QY 441 AsnGluGlyTrpSerSerLeuGlnArgSerAlaHisLysLeuAsnGluThrGlyAlaAsp 460  
 DB 1433 AATGAAGCTGGGCCAATATAGAGAGTGTGCTCAACTGCTCTCAAGAGACAGTCAAGT 1492  
 QY 461 PheIleThrIleLeuGlnSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThr 480  
 DB 1493 TTTATCACCATTGTTGGAGAGCATGCTTCCAAAGCCCTATATAGGAGAACACACTTAACA 1552  
 QY 481 MetTrpLeuGlyGlnLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHis 500  
 DB 1553 ATGTGGCTGGGGAGAACTGGGTTTCTACACAGACTTTGGGCCAAGCCAGGATGAC 1612  
 QY 501 ThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGlnHisHisLeuLeu 520  
 DB 1613 ACCTGGGAGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672  
 QY 521 ProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnLysSerGlyLys 540  
 DB 1673 CCGTGGCCAGAGGCGAGATCCACACAGCATACCATATGACATGACATGCTCCAAACAGA 1732  
 QY 541 LeuValAspPheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeu 560  
 DB 1733 CTGGTGGATTTTGTGTGACACACTTTGGGAATCAAGAAAGACCTTGGACGGAAGCTA 1792  
 QY 561 GlnAlaIleAlaValSerLysLeuLeuLysSerSerSerSerSerGlnValIlePheLeuGly 580  
 DB 1793 CAGGCTATTTGCAAGTTTCAAAACTGCTGAAGAAATTTGTCAAACTGATATTTCTGGGT 1852  
 QY 581 TyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThrGlnHisGlyAsnVal 600  
 DB 1853 TATATCATTCTGACGACTCGTGTCCAGAGATTACAAACACTCACAACACTGGCAACGTG 1912  
 QY 601 LysAspIleAspSerThrAspHisAspArgTyrCysGluTyrIleMetTyrArgGlyLeu 620  
 DB 1913 AAGCATATTGACAGCTCAGATGAGACAGAGTGTGTAGTACATATGATACGGGGCTGTG 1972  
 QY 621 IleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMet 640  
 DB 1973 ATCAGATGGGCTATGCAAGAACTCTGATGCAAGATGAGTACGATGATGATGATGATGATG 2032  
 QY 641 AlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGlnLysValIleIle 660  
 DB 2033 GCCAAATTCAGGATCCGATATATCCGCCAATTTACAGAGACACAAACACTGCTCATTA 2092  
 QY 661 AspHisArgGluValSerGluLysLysIleHisPheAsnProArgPheGlySerTyrLysGlu 680  
 DB 2093 GACCCACCGAGAGTTCCCAAGAACATTTCACTTCAACCCCAAGATTTGGCTCTACAAAGAT 2152  
 QY 681 GlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPheLeu 699  
 DB 2153 GGACACAATTTAGAGAACACCACATCATTTTTCATGATGATACCCCAATTAATCTGCTT 2209  
 RESULT 3  
 SCCHRII/c 316613 bp DNA linear PLN 29-JUN-2001  
 LOCUS S\_cerevisiae chromosome III complete DNA sequence.  
 DEFINITION X59720.2 S43845 S49180 S58084 S93798  
 ACCESSION X59720.2 GI:14588895  
 VERSION 1  
 KEYWORDS Chromosome.  
 SOURCE Baker's yeast.  
 ORGANISM Saccharomyces cerevisiae  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
 REFERENCE 1 (bases 1 to 316613)  
 AUTHORS Rad,M.R., Lutzenkirchen,K., Xu,G., Kleinhaus,U. and Hollenberg,C.P.  
 TITLE The complete sequence of a 11,953 bp fragment from Clg on



JOURNAL MEDLINE REFERENCE AUTHORS TITLE  
chromosome III encompasses four new open reading frames  
Yeast 7 (5), 533-538 (1991)  
91377317  
2 (bases 1 to 316613)  
Wilson, C., Bergantino, E., Lanfranchi, G., Valle, G., Carignani, G. and Frontali, L.  
A putative serine/threonine protein kinase gene on chromosome III of *Saccharomyces cerevisiae*  
Yeast 8 (1), 71-77 (1992)  
92254506  
3 (bases 1 to 316613)  
Biteau, N., Fremaux, C., Hebrard, S., Menara, A., Aigle, M. and Crouzet, M.  
The complete sequence of a 10.8kb fragment to the right of the chromosome III centromere of *Saccharomyces cerevisiae*  
Yeast 8 (1), 61-70 (1992)  
92254505  
4 (bases 1 to 316613)  
Benit, P., Chanet, R., Fabre, F., Faye, G., Fukuhara, H. and Sor, F.  
Sequence of the sup61-RAD18 region on chromosome III of *Saccharomyces cerevisiae*  
Yeast 8 (2), 147-153 (1992)  
92221691  
5 (bases 1 to 316613)  
Bolle, P.A., Gilliquet, V., Berben, G., Dumont, J. and Hilger, F.  
The complete sequence of K3B, a 7.9 kb fragment between PGK1 and CRY1 on chromosome III, reveals the presence of seven open reading frames  
Yeast 8 (3), 205-213 (1992)  
92245758  
6 (bases 1 to 316613)  
Sor, F., Cheret, G., Fabre, F., Faye, G. and Fukuhara, H.  
Sequence of the HMR region on chromosome III of *Saccharomyces cerevisiae*  
Yeast 8 (3), 215-222 (1992)  
92245759  
7 (bases 1 to 316613)  
Skala, J., Purnelle, B. and Goffeau, A.  
The complete sequence of a 10.8 kb segment distal of SUP2 on the right arm of chromosome III from *Saccharomyces cerevisiae* reveals seven open reading frames including the RVS161, ADP1 and PGK genes  
Yeast 8 (5), 409-417 (1992)  
92327849  
8 (bases 1 to 316613)  
Oliver, S.G., van der Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P., Benit, P. et al.  
The complete DNA sequence of yeast chromosome III  
Nature 357 (6373), 38-46 (1992)  
92244356  
9 (bases 1 to 316613)  
Scherens, B., Messenguy, F., Gigot, D. and Dubois, E.  
The complete sequence of a 9,543 bp segment on the left arm of chromosome III reveals five open reading frames including glucokinase and the protein disulfide isomerase  
Yeast 8 (7), 577-585 (1992)  
92397595  
10 (bases 1 to 316613)  
Wilson, C., Grisanti, P. and Frontali, L.  
The complete sequence of a 6146 bp fragment of *Saccharomyces cerevisiae* chromosome III contains two new open reading frames  
Yeast 8 (7), 569-575 (1992)  
92397594  
11 (bases 26740 to 32076)  
Defoot, E., Debrabant, R., Keyers, B., Voet, M. and Volckaert, G.  
Nucleotide sequence of D10B, a BamHI fragment on the small-ring chromosome III of *Saccharomyces cerevisiae*  
Yeast 8 (8), 681-687 (1992)  
93070606  
12 (bases 169581 to 171116; 171683 to 172169)  
Agostoni-Carbone, M.L., Panzeri, L., Muzi Falconi, M., Carcano, C., Plevani, P. and Lucchini, G.  
Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III

JOURNAL MEDLINE REFERENCE AUTHORS TITLE  
from strain AB972: evidence for a Ty insertion and functional analysis of open reading frame YCR28  
Yeast 8 (9), 805-812 (1992)  
93070619  
13 (bases 1 to 315338)  
Valle, G.  
TA-repeat microsatellites are closely associated with ARS consensus sequences in yeast chromosome III  
Yeast 9 (7), 753-759 (1993)  
93377412  
14 (bases 1 to 315338)  
Slonimski, P. and Brouillet, S.  
A data-base of chromosome III of *Saccharomyces cerevisiae*  
Yeast 9 (9), 941-1029 (1993)  
94091061  
15 (bases 1 to 315339)  
Rodriguez-Cousino, N., Lill, R., Neupert, W. and Court, D.A.  
Identification and initial characterization of the cytosolic protein Ycr77p  
Yeast 11 (6), 581-585 (1995)  
95373282  
16 (bases 1 to 316613)  
MIPS.  
Direct Submission  
Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing Project  
replaced by [13]  
17 (bases 1 to 314957)  
Jimenez, A.  
Direct Submission  
Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing Project. Update originating from A. Jimenez  
replaced by [14]  
18 (bases 1 to 316613)  
Louis, E.J.  
Direct Submission  
Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing Project. Update originating from E.J. Louis  
revised by [18]  
19 (bases 1 to 316613)  
Louis, E.J.  
Direct Submission  
Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing Project. Update originating from E.J. Louis  
revised by [20]  
20 (bases 1 to 316613)  
Gromadka, R.  
Direct Submission  
Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis, Institute of Biochemistry and Biophysics, Pawlinski 5A, Warsaw, Poland, Electronic Mail Address: robertpsd.ldb.waw.pl  
revised by [21]  
21 (bases 1 to 316613)  
MIPS.  
Direct Submission  
Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF - Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany, mips-yeast-adm@gsf.de  
Resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. Munich Information center for Protein Sequences.  
On Jul 3, 2001 this sequence version replaced g1:1907116.  
The contig sequence which has been released to the file server in 1992 has been subject to a resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. The following lines present a summary of the altered entities. The resequenced chrIII contig is 316613 bp with GCG-check: 429 GENETIC ENTITIES valid no longer:  
YCL012W (YCL014W + YCL012W -> YCL014W / C\_B1367 + C\_C231->C\_A1636),



```

Db 145612 CCTATATCATCCAGATTCTACGTTGATTCACGCTGATTTGACTATCCATCTTGGTCTC 145553
OY 440 ASPANGLUGLYTRPSESRLEUENLARGSERIALHISLEULANSGLUHRGALYA 459
Db 145552 GATTAATGATAGTGGCATCTGAAGACAGATGATCACTTATTAATAATVGAACATA 145493
OY 460 ASPHEILIEHRILELEUGLUSERASPAISERLYSPROTYMETGLYASNAASPLEU 479
Db 145492 GATGGTAGAGTCTACTAGAAACATACACAAAGATTAACATGGGAGGAGGATCTCA 145433
OY 480 THRMEETRPLEUGLYLULYSLEUGLYPHERYTRHASPHEGLYPROSETHRAIRGYR 499
Db 145432 ACTACCAACACTAGCTCATGATTTGAATATGATGACAGATTTGGACACAGTCAATAA 145373
OY 500 HSTHTRPSGLYILEMETALALEUSERARGYRPROLIEVALYLSERLUNHSHISLEU 519
Db 145372 CATACCTGGGGCTGTGCTCTTCTTCTTAATTCCTATCCATAATTCATGCGATCATTTTA 145313
OY 520 LEUPROSERPROGLUGLYLULLEALAPROLAILETHLEUTHRVAL--ASNILESER 538
Db 145312 TTGGCCCTCTCCAGTGGGGAACCTTGGCCAGCCATTCATGCCACTTCAACGTAACAT 145253
OY 539 GLYLSLEUALASPHEVALVALTHRHSPEGLYASNHISGLUASPAISLEUASPAR 558
Db 145252 GACACTCTGCTGACGCTCTTGTATTCATAGTGAACAAGAGAGATGACAGAGATAGA 145193
OY 559 LYSLEUGINALAILEALVALSERLYSLEULEULYSSESRSERASNGINVALILEPHE 578
Db 145192 AGACGCTGAAGAACTACATGCTTAAGCTCATGGCAATACGATCCGACCTATTTTA 145133
OY 579 LEUGLYTRYLIEHRSERALARPROGLYSERARGASPTYR--LEUGLULEUTHRGUHS 597
Db 145132 TTAAGTACTAGTGTGTTGATCCAGCTGAAGCACTACCAATACGTAACCTGATGTAACA 145073
OY 598 GLYASNAVLYSASPILEASPSERTHRASPHISAPARGTPCYSGLUYRILEMETYR 617
Db 145072 TCCGGAATGACGACATGATGCTCTCTGACGATGATAGATGATGATATCTGTAT 145013
OY 618 ARGGLYLEULLEARGLEUGLYTRYRALARGILESERHISALGLULEUSERASPSERGL 637
Db 145012 AGGGGCTTGAGAAAGAACAGGATATGCTAGAGTTCAGAGAGAACGATACCATACGAG 144953
OY 638 ILEGIMETALALYSRPHETRGILEPROASPARPROTHASNTYRARGSPASNGINL 657
Db 144952 CTACAAAGTTGGTAAGTTCACAACTTTTG-----AGTGAGCA 144917
OY 658 VALVALILEASPHISARGGLU-----VALSERGLULYSILEHIS 670
Db 144916 GCGTAGTAGACACTCGGATTCATATGATGAATACGTCATATGATGACCGGAATAT 144857
OY 671 PHEASNPORARGPHEGLYSERYTRYLSGLUGLYHISASNTYRGLUASNAHSHISPLE 690
Db 144856 GAGGACATGAATAATTCAGATTAAGTTTAAAGCGAAGT--GAGAGGGGCTCATCTTCTAC 144800
OY 691 HISMET--ASNTHRPROLYSTYRPHLEU 699
Db 144799 CATGTTTGTGATGACGACGTTATTACTTA 144770

```

```

RESULT 4
SPAC688 34034 bp DNA linear PLN 08-MAY-2000
LOCUS S.pombe chromosome I cosmid c688.
DEFINITION AL355632.1 GI:7768476
ACCESSION
VERSION
KEYWORDS

```

Actin-related protein; cell polarization; coiled-coil; cut9  
 interacting protein; dna polymerase zeta catalytic subunit; drug  
 sensitivity; DUF61; endocytosis internalization phase; ENTH domain;  
 glutathione S transferase; I/LWDQ domain; korberg's mediator (SRB)  
 subcomplex; membrane cytoskeleton assembly; microtubule stability;  
 mitochondrial carrier protein; Protein of unknown function domain;  
 RNA polymerase II holoenzyme; scrl; Tfi-type LTR; yeast Sla2; yeast  
 SRB8; yeast YCR017C.

**SOURCE**  
 ORGANISM  
 Schizosaccharomyces pombe  
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 Schizosaccharomycetales; Schizosaccharomycetaceae;  
 Schizosaccharomycetes  
 1 (bases 1 to 34034)  
**REFERENCE**  
 AUTHORS  
 Aert, R., Robben, J., Volckaert, G., Wood, V., Rajandream, M.A. and  
 Barrell, B.G.  
**TITLE**  
 JOURNAL  
 Direct Submission  
 Submitted (15-APR-2000) European Schizosaccharomycetes genome  
 sequencing project, Sanger Centre, The Wellcome Trust Genome  
 Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk  
 and Katholieke Universiteit Leuven, Laboratory of Gene Technology,  
 Kardinaal Mercierlaan 92, B-3001 Leuven, Belgium  
**NOTES**

**COMMENT**  
 Details of yeast sequencing at the Sanger Centre are available on  
 the World Wide Web.  
 (URL, <http://www.sanger.ac.uk/Projects/S.pombe/>)  
 During 1995 to 1996 about 66% of S. pombe chromosome I was  
 sequenced by the Sanger Centre. The sequencing of the S. pombe  
 genome is now being continued with funding from the European  
 Commission. Fourteen European sequencing laboratories, including  
 the Sanger Centre, are participating in the project.  
 Protein coding regions (CDS) have been predicted with the help of  
 computer analysis using the GeneFINDER program in Pombase (an ACEDB  
 database) with additional predictions for the branch-acceptor sites  
 supplied by the program Sp3splice. CAUTION: It is possible that for  
 any individual CDS we may have underestimated or overestimated the  
 number of introns/exons or we may not have chosen the correct  
 splice donor/acceptor sites.  
 CDS are numbered using the following system eg SPAC25H2.01c. SP (S.  
 pombe), B (chromosome 2), c25h2 (cosmid name), .01 (first CDS), c  
 (complementary strand).  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
**IMPORTANT:** This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions.  
 Cosmid c688 is overlapped at the 3' end by cosmid c369, EMBL entry  
 SPAC369, accession number AL021046.  
**FEATURES**  
 source  
 location/Qualifiers  
 1. 34034  
 /organism="Schizosaccharomycetes pombe"  
 /strain="972h"  
 /db\_xref="taxon:4896"  
 /chromosome="I"  
 /map="IL"  
 /clone="cosmid c688"  
 1. 1280  
 /gene="SPAC688.01"  
 1. 1280  
 /partial  
 /gene="SPAC688.01"  
 /note="SPAC688.01, len:425, SIMILARITY:Saccharomycetes  
 cerevisiae, YC07\_YEAST, hypothetical 107.9 kd protein,  
 (953 aa), fasta scores: opt: 1034, E():0, (46.3% identity  
 in 430 aa)"  
 /codon\_start=3  
 /label=SPAC688.01  
 /product="hypothetical protein similar to yeast YCR017C  
 protein involved in sensitivity to certain drugs"  
 /protein\_id="CAB90768.1"  
 /db\_xref="GI:7768477"  
 /translation="SYFMVINGMCSYLGILFISYVLIYSFASIRISSPYSPARKWG  
 GAFVLIYLSLAHVVAVFVPGGPIILERTSYLIFIGMLALVAVSGESKEPN  
 KADSSVVIDIKOSDSYRRSRKSLITGFLALMLAKRAIONMPYDVTTPHPNKKPL  
 TAGITHTFGIDNPYASENRTIRDAVRMEDYRGLLESIDPQRILMGRODLYOVLAND  
 LGMATDYGPKHTWGAALSKRPFTVSTHLLPSPGELAPATHATLDVYGGELIDV  
 VSHNGQYESOLDRLRLOSTELARIMRESPRLVFGVYAVNGDEPQITLRDGMID  
 IEPADYDMCOYIIFRGVARKIGYALRHSTLTDELTEGKFLVTKDIGNRVRIKKEHV  
 PESHRYPLSEGTGVNGHYDNNLVHBPWYVD"



```

Oy 367 -----GlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLys 379
Db 279 GCGATTCGGGAGTCCAAAGAACCAAGACAGACTCATCTGTAGTATATCAAA 338
Oy 380 AsnSerSerLysValLeuPheArgLysSerGlyLysTyrMetLysLeuPheLeuTrpLeu 399
Db 339 CAATCAGATTTCTTCTTATCTAGAGAGAGTTCACAAATAATCA-----CTA 383
Oy 400 LeuValGlyValGlyLeuLeuGlyLeuGlyLeuGlnHisLysAlaTyrGluArgLysLeu 419
Db 384 TTGACAGGTTTGTGTCTGCTTATGAGCCCTTAATAATTTGCCATTGCAGAAATATGCTCCG 443
Oy 420 GlyValAlaPro-----ThrLysGluValSerAlaAlaIleTrpProPhe 435
Db 444 TATGACTTACACCTTACCATCCAAATGAAAAGCTTTCACGCTGGATTTGGACTATT 503
Oy 436 ArgPheGlyTyrAspAsnGluGlyTyrSerSerLeuGluArgSerAlaHisLeuAsn 455
Db 504 CACTTGGCTTAGACAAATTCATGTATGCAAGTGAAGTAACTGATACGGCATGCTGTGCA 563
Oy 456 GluThrGlyAlaAspPheIleThrLeuGlnSerAlaSerLysProGlyMetGly 475
Db 564 GATATGACACTGTGATGCTGTGCTGCTGATCTGACACACACAGCTTAAATTTGGGA 623
Oy 476 AsnAsnAspLeuThrMetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyPro 495
Db 624 TTTCAGATTTAACCCAGCTTTTACACATGATCTAGAGAAATGTATAGCGATATATGCCCA 683
Oy 496 SerThrArgTyrHisThrTrpGlyLeuMetAlaLeuSerArgTyrProIleValLysSer 515
Db 684 GGCCCTGATTAACATATCTGGGCGCTGCTGATCTGCAATTCCTATGTCATGTCACACTCG 743
Oy 516 GlnHisHisLeuLeuProSerProGluGlyGluIleAlaProAlaIleThrLeuThrVal 535
Db 744 ACTGACCACTTGTATCTCTCTCCACAGAGAACTTGTCTCTGCTATTCATGACACACTA 803
Oy 536 AsnIleSerGlyLysLeuValAspPheValValThrHisPheGlyAsnHisGluAspAsp 555
Db 804 GATGCTACGGAGACTATAGATCTGTGCTTCTCACAAATGTCATATCAAAAGTCCAG 863
Oy 556 LeuAspArgLysLeuGlnAlaIleAlaValSerLysLeuLeuLysSerSerAsnGln 575
Db 864 TTGGATTCACAGCTTCACAGACACTGAGTGGCTCGAATATGCGCAGAGTCCAAAGCCT 923
Oy 576 ValIlePheLeuGlyTyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThr 595
Db 924 CTAGTGTCTCTCGATATGTTGTTCAACGTTGGCCCAAGAACCCCAAACTATTTCTAACG 983
Oy 596 GlnHisGlyAsnValLysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIle 615
Db 984 CGAGACACGGGAATGTTAGACATTTGACCTGCTGACTATGATGATGTTGGTTCAGTATATA 1043
Oy 616 MetTyrArgGlyLeuIleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAsp 635
Db 1044 TTTTATCGAGGTGTACAGAGCAATGTGTTATGCTAGACTTCATCAGCTCTACTATACAGAC 1103
Oy 636 SerGluIleGlnMetAlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsn 655
Db 1104 ACCGAGCTTCAACCCGGAAGATTTTGGTTACCAAAAGTTTA-----GGT 1148
Oy 656 GlnLysValValIleAspHisArgGluValSerGluLysIleHisPheAsnProArgPhe 675
Db 1149 AGGACGTCAGATGACAGACAGACAGCTTCTGATGATCAGAGATATCATCATTTATT 1208
Oy 676 ---GlySerTyrLysGluGlyHisAsnTyrGluAsnAsnHisPheHisMetAsnThr 694
Db 1209 GAAGGACGGGTGTAAATGACATTAATGACAAATAATATTAGTTGTTAC-----GAG 1262
Oy 695 ProLysTyrPhe 698
Db 1263 CCGGTGTAATA 1274

```

RESULT 5

```

D89130
LOCUS D89130 1234 bp mRNA linear PLN 13-MAR-1998
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0503.
ACCESSION D89130
VERSION D89130.1 GI:1749467
KEYWORDS
SOURCE
ORGANISM
Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA,
clone.lib:library of H. Nojima clone:SY 0503.
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 1234)
yoshioaka,S.
REFERENCE
AUTHORS Yoshioaka,S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1996) Sachio Yoshioaka, Tsukita Cell Axis Project
ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi,
Shimokyo-ku, Kyoto, Kyoto 600, Japan
(E-mail: syoshi@cell.tsukita.jst.go.jp, Tel:+81-75-7913,
Fax:+81-75-315-6420)
2 (sites)
REFERENCE
AUTHORS Yoshioaka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
TITLE Identification of open reading frames in Schizosaccharomyces pombe
cDNAs
JOURNAL DNA Res. 4 (6), 363-369 (1997)
FEATURES
source
location/Qualifiers
1..1234
/organism="Schizosaccharomyces pombe"
/strain="PR745"
/db_xref="taxon:4896"
/clone="SY 0503"
/clone.lib="library of H. Nojima"
<1..1207
/Note="Similar to Saccharomyces cerevisiae hypothetical
107.9KD protein in POL-4-SRI intergenic region, SWISS-PROT
Accession Number P25618"
/codon_start=2
/protein_id="BAJ3792.1"
/db_xref="GI:1749468"
/translacion="TSPASIRISSFYSAPKAWGAGLVLYLSIAHVVAVEFPV
GPILRERTSYLILFIGNLALVPAVSESEKPKADSVVDIKQSDSYRRSEKKS
LITGFCALMALKFAIOMPPDYTPYHPNKELETTAGITLHFGIDNPMYASENIR
AVRWELDFVGLSDFORLIMGFDLVOVLAHDGMYADVGPGRDKHTGWAALLSK
PIVNSTHLLPSQGEELAPAIHATIDYGEILIDVVSINQGEESQIDRLQSTELAI
MRESPRLVFLGIYVSNVNGQEPQITLITDQTMLEIDPADYRDKOYITRVKAKIGIA
RLHNSITDTELQTEKFLVTKDLGNVRIDKEHVPESHRYPSLEFGTGVNGHYDNNL
VVEHPWYVD"
BASE COUNT 333 a 253 c 253 g 395 t
ORIGIN
Alignment Scores:
Pred. No.: 4.16e-42 Length: 1234
Score: 625.50 Matches: 145
Percent Similarity: 50.98% Conservative: 63
Best Local Similarity: 35.54% Mismatches: 169
Query Match: 16.81% Indels: 31
Gaps: 7
US-09-809-638-2 (1-699) x D89130 (1-1234)
Oy 306 IleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeu 325
Db 26 ATTTCCTCTTATTTCTCTCTCGAGAGTTTGGGGTGGGCGCTTTTGGTTTACATCTT 85
Oy 326 GluIlePhePheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAla 345
Db 86 TACTCATTTAGTCATCTTTGGGTCGTAAGCTTCAAGATTTGTTCCAGGTGAGCCATTTCT 145
Oy 346 ArgGluArgSerAspValLeuGlyThrMetMetLeuIleGlyLeuAsnMetLeu 365
Db 146 CGAGAGAACTTCTTACATTTCTA-----ATTTCATTTGGTTGGAACCTTGCA 193

```

```

Oy 366 phe-----GlyProLysLysAsnLeuLeu 375
Db 194 GCTTGTAGTACGCGTATTGCGGTGAGTCCAAAGAACACCAACAGACTCATCTGTA 253
Oy 376 LeuGlnThrLysAsnSerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeu 395
Db 254 GTAGATATCAACAACAAATCTTCTTATCGTAGAAGAGTTTCAAAAATCA----- 307
Oy 396 PheLeuTrpLeuValGlyValGlyLeuGlyLeuArgHisLysAlaTyr 415
Db 308 -----CTATTGACAGGTTTCTGCTGCTTATGGCCCTAAATTTGGCATTCAG 358
Oy 416 GluArgLysLeuGlyLysValAlaPro-----ThrLysGluValSerAlaAla 431
Db 355 AATATGCTCCGTATGACTATACACTTACCATCAATGAATAAGACTTTTCACTGGT 418
Oy 432 IleTrpProPheArgPheGlyTyrAspAsnGluGlyTrpSerSerLeuGluArgSerAla 451
Db 419 ATTTGGACTATTTCACCTTGGCTTAGACAATTTTCATGTATGCAAGATGAGATATACGC 478
Oy 452 HisLeuLeuAsnGluThrGlyAlaAspPheIleThrIleLeuGluSerAlaSerLys 471
Db 479 GATGCGTTCGAGATATGGAAGTGTATGCTTGTCTACTGGAATCTGACACACACAGT 538
Oy 472 ProTyrMetGlyAsnAsnAspLeuThrMetTrpLeuGlyGluLysLeuGlyPheTyrThr 491
Db 539 TTAATATTGGATTTCGAGATTTAAACCCAGTTTATGACACATGATCTAGGAATGTATGCG 598
Oy 492 AspPheGlyProSerThrArgTyrHisThrTrpGlyIleMetAlaLeuSerArgTyrPro 511
Db 599 GATTATGGCCAGGCGCTGATTAACACTTGGGCGCTGCGCTTACTCTCCAAATTTCCCT 658
Oy 512 IleValLysSerGluHisLeuLeuProSerProGluGlyGluIleAlaProAlaIle 531
Db 659 ATTGTCACTGACATCCACATTTGTTACTCTTCCCTCAAGACAACTGCTCTCTATT 718
Oy 532 ThrLeuThrValAsnIleSerGlyLysLeuValAspPheValValThrHisPheGlyAsn 551
Db 719 CATGCAACTAGATGTCTACGAGAGCTAATAGATGTGTGTTCTCCAAATGCTCA 778
Oy 552 HisGluAspAspLeuAspArgLysLeuGlnAlaIleAlaValSerLysLeuLeuSer 571
Db 779 TATGAAAGTCACTGATGATGAGCTCTTCAGACACTGAGTGGCTCGAATTTATGCGCAG 838
Oy 572 SerSerAsnGluValIlePheLeuGlyTyrIleThrSerAlaProGlySerArgAspTyr 591
Db 839 ACTCCAGGCCCTACTGTTCTCGATATGTTGTTTCAACGTTGCCAAGACCCCA 898
Oy 592 LeuGlnLeuThrGluHisGlyAsnValLysAspIleAspSerThrAspHisAspArgTyr 611
Db 899 ACTATTCTTAACCGACAGACACGGAAGTTAGACATGTAACCTGCTGACTATGATCGTGG 958
Oy 612 CysGluTyrIleMetLysTrpArgLysLeuIleArgLeuGlyTyrAlaArgIleSerHisAla 631
Db 959 TGTCACTATATTTTATTTAGAGTGTCAAGCGAATTTGTTATGCTTATATCGCTCT 1018
Oy 632 GluLeuSerAspSerGluIleGlnMetAlaLysPheArgIleProAspAspProThrAsn 651
Db 1019 ACTATAACAGACACCGACTTCAACCGGAAGTTTGGTTACCAAGATTTA----- 1072
Oy 652 TyrArgAspAsnGlnLysValIleAspHisArgGluValSerGluLysIleHisPhe 671
Db 1073 -----GTTAGAGAACTCAGAAATCGACAAAGAACACAGCTTCCTGATCACAGAGTAT 1123
Oy 672 AsnProAlaGpHe---GlySerTyrLysGluGlyHisAsnTyrGluAsnHisHisPhe 690
Db 1124 CCATCATATTTGAAGAACGGGTGTAAATGACATTTACTATGACAAATATTTAGTTGTT 1183
Oy 691 HisMetAsnThrProLysTyrPhe 698
Db 1184 CAC-----GAGCCGTGTTACTAT 1201

```

```

CNS06JWV
LOCUS      995 bp      DNA      linear      STS 10 JAN-2001
DEFINITION T3 end of clone XAS0A002B08 of library XAS0A from strain CLIB 533
ACCESSION AL402149
VERSION    AL402149.1 GI:12160780
KEYWORDS   Saccharomyces bayanus.
SOURCE     Saccharomyces bayanus.
ORGANISM   Saccharomyces bayanus.
REFERENCE  1 (bases 1 to 995)
AUTHORS    Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
           Aigle,M. and Durrens,P.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 5.
           Saccharomyces bayanus var. uvarum
JOURNAL    FEBS Lett. 487 (1), 37-41 (2000)
PUBMED     11152860
AUTHORS    2 (bases 1 to 995)
           Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
           Boloitin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
           de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
           Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
           Sautin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
           Wincker,P. and Weissbach,J.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
           yeast species for molecular evolution studies(1)
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
PUBMED     11152876
AUTHORS    3 (bases 1 to 995)
           Direct Submission
           Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
           2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
           seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT    This STS is part of a random genomic sequencing program of thirteen
           yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
           exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
           Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
           lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
           angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
           Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
           5 kb were prepared and both extremities were sequenced. See
           keywords for description of this sequence and for the sequence of
           the other extremity of this insert.
FEATURES   Location/Qualifiers
           source          1..995
                        /organism="Saccharomyces bayanus"
                        /strain="CLIB 533"
                        /variety="uvarum"
                        /db_xref="taxon:4931"
                        /clone="XAS0A002B08"
                        /clone_1lb="XAS0A"
                        /note="end : T3"
           misc-feature    <3..>998
                        /note="similar to Saccharomyces cerevisiae ORF YGR017c [
                        similarity to hypothetical S.pombe protein ]
                        putative frameshift(s)"
           BASE COUNT     276 a      190 c      210 g      318 t      1 others
           ORIGIN
Alignment Scores:
Pred. No.:      1..53e-31      Length:      995
Score:          495.00          Matches:      136
Percent Similarity: 50.55%      Conservative: 48
Best Local Similarity: 37.36%      Mismatches:  136
Query Match:    13.31%          Indels:      47
DB:             11              Gaps:        10

```

Oy 274 GlyLeuLeuTyrLeuHisThrTrpAlaAlaValAlaSerGlyCysValPheAlaIlePhe 293

```

Db      6 GGTATTATGTAC-----GCGATGTGCTATGCTTTGGTGTAGTCTCCTCATACTTT 53
Oy      294 ThrAla-----SerMetTrpProGlnThrLeuGlyHisLeuIleAsnSer 308
Db      54 TCCGACTAGGTGCAAGTACAAAGATATATGCGTTTACGTTTTC-----95
Oy      309 GlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGluIlePhe 328
Db      96 -----TCATTTCCTGCTATGTTATTTTGTGCTCATCTTT---134
Oy      329 PheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArgGluArg 348
Db      135 -----TGGGTGCTGCAATATGCATTCCTTCATAGGTGGTTACTGAGGAGAGA 185
Oy      349 SerAspValLeuLeuGlyThrMetLeuIleIleGlyLeuAsnMetLeuPheGlyPro 368
Db      186 ATCGAGGCTATCTCCGAGTTTCTCGACGTTTATTATCTTGATGTGTCACAAAT 245
Oy      369 LysLysAsnLeuAspLeuLeuLeuGlnThrLysAsnSerLysValLeuPheArgLys 388
Db      246 AACCTCAACCTTCGACTATA-----AACGTAGCAGAAATTTATGC-----287
Oy      389 SerGlyLysTyrMetLysLeuPheLeuTyrLeuValGlyValGlyLeuGlyLeu 408
Db      288 -----ACTATATTTCTTTGTT-----GCGATAGCTCTACTAGCATTCG 326
Oy      409 GlyLeuArgHisLysAlaTyrGluArgLysLeuGlyLysValAlaPro-----424
Db      327 ACCGCTAGATTACATAGCAATAGACCACCGGAATTCGCCACCTATCATCTCGAT 386
Oy      425 ThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAspAsnGluLysTrp 444
Db      387 TCTCATATGATCTGCTGGTATTTGGACCATTCATTGTTGCTGATATATATATGTCG 446
Oy      445 SerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAspPheIleThrIle 464
Db      447 GCATGTGAAGACAGATGATCACTTAATAAGAAATGCAATGATGATTTAGATTA 506
Oy      465 LeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThrMetTrpLeuGly 484
Db      507 CTAGAAACAGATACCAAAATTTATTTGGTAAACAGGATTAACCACTAATCAATGCC 566
Oy      485 GluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHisThrTrpGlyLe 504
Db      567 CATGATCTGAATATGATGCTGATTTTGGCCCGGCTCCAAACAAACACACATGGGGCTGT 626
Oy      505 MetalLeuSerArgTyrProIleValLysSerGluHisLeuLeuProSerProGlu 524
Db      627 ATCTTCTTCCAAATCCCTATCATTAATGCAACGACCATCTATGCGATCCCTGTG 686
Oy      525 GlyGluIleAlaProAlaIleThrLeuThrVal---AsnIleSerGlyLysLeuValAsp 543
Db      687 GGAGAGCTTGGCCCGCCGCTATCCGCCACTCTTCAACGCTAATATGACACGCTAGTGTAT 746
Oy      544 PheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeuIleAlaIle 563
Db      747 GCTTGTGCTTTCATATGAGGACAGAGGAAAGATGAAGAAATGAATTAACAAGTAAAT 806
Oy      564 AlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePheLeuGlyTyrIleThr 583
Db      807 TACATGGCTAATTTGATGGAAATTCGACCCGACACGATTTTATTAAGTACTTGGTG 866
Oy      584 SerAlaProGlySerArgAspTyr---LeuGlnLeuThrGluHisGlyAsnValLysAsp 602
Db      867 GTTGACCCGCGGAGGAACTACATACGATGTAGTGAAGT---TCTGGCATGTCATGAC 925
Oy      603 IleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeuLeuArg 622
Db      926 ATTGACCCCTACTGACGATGATAGTGGTGTGA---TATATCTGTATTAAGATTGAGACAGA 984
Oy      623 LeuGlyTyrAla 626

```

```

Db      985 CA-GGKTATGCA 995
RESULT 7
LOCUS      AC020593
DEFINITION Homo sapiens BAC clone RP11-317G22 from 4, complete sequence.
ACCESSION AC020593
VERSION    AC020593.6
KEYWORDS   GI:14589709
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 184206)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 184206)
Swearengen, S. and Cotton, M.
The sequence of Homo sapiens BAC clone RP11-317G22
Unpublished
3 (bases 1 to 184206)
Waterston, R.H.
Direct Submission
Submitted (05-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 184206)
Waterston, R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 184206)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 3, 2001 this sequence version replaced gi:12057004.
-----
Genome Center
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
-----
Summary Statistics
Center project name: H_NH0317622
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBAC3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is AC011956. Actual start of this clone is at base position 1 of RP11-317G22; actual end is at base position 184206 of RP11-317G22.

The sequence fidelity of RP11-317G22 between bases 17725 to 17956 can not be guaranteed due to an unresolved dinucleotide repeat. The sequence fidelity of RP11-317G22 between bases 86965 to 87015 can not be guaranteed due to an unresolved homopolymeric run.

## FEATURES

```

source
1..184206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-317G22"
/clone_lib="RPCI-11"
5..70
/rpc_family="L2"
misc_feature
252..374
/note="match to EST BG169511 (NID:g12676214)"
misc_feature
252..374
/note="match to EST AI088656 (NID:g3427715) qbl4a11.x1"
misc_feature
252..374
/note="match to EST BG214746 (NID:g13740767)"
misc_feature
252..374
/note="match to EST BF055447 (NID:g10809343)"
misc_feature
252..374
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
misc_feature
252..374
/note="match to EST AM006753 (NID:g5855531) wt07a02.x1"
misc_feature
252..374
/note="match to EST BG291592 (NID:g13049843)"
misc_feature
252..374
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature
252..494
/note="match to EST AA464688 (NID:g2189572) zx82a02.s1"
misc_feature
252..374
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
misc_feature
252..374
/note="match to EST BG165990 (NID:g12672693)"
misc_feature
256..374
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
280..374
/note="match to EST AW371794 (NID:g6876448)"
347..374
/note="match to EST AA303948 (NID:g1956281)"
708..1203
/note="match to EST BI063872 (NID:g14471399)"
1021..1308
/note="match to EST T06930 (NID:g318079)"
1119..1143
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
1119..1143
/note="match to EST AI220135 (NID:g3802338) q144g05.x1"
1119..1143
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
1183..1538
/note="match to EST BG167015 (NID:g12673728)"
1183..1538
/note="match to EST BG546183 (NID:g13544848)"
1319..1350
/rpc_family="AT_rich"
1349..1466
/rpc_family="Alu"
1490..1538
/note="match to EST AA847380 (NID:g2933898) of07d05.s1"
1490..1538
misc_feature

```

```

/note="match to EST AI220135 (NID:g3802338) q144g05.x1"
1490..1538
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
1490..1538
/note="match to EST BG167015 (NID:g12673728)"
1490..1538
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
1491..1538
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
1491..1538
/note="match to EST AI088656 (NID:g3427715) qbl4a11.x1"
1491..1508
/note="match to EST BG291592 (NID:g13049843)"
1491..1538
/note="similar to Homo sapiens EST BG207366 (NID:g13729053)"
1491..1538
/note="match to EST BG165990 (NID:g12672693)"
1491..1538
/note="match to EST AA303948 (NID:g1956281)"
1491..1538
/note="match to EST AW371794 (NID:g6876448)"
1491..1538
/note="match to EST AM005009 (NID:g5853713) wv02e08.x1"
1491..1538
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
1491..1538
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
1491..1538
/note="match to EST BG214746 (NID:g13740767)"
1491..1538
/note="match to EST AM006753 (NID:g5855531) wt07a02.x1"
1491..1538
/note="match to EST BG169511 (NID:g12676214)"
1491..1546
/note="match to EST BF055447 (NID:g10809343)"
1561..1606
/rpc_family="AT_rich"
1696..1986
/rpc_family="Alu"
2175..2217
/rpc_family="Alu"
2311..2780
/rpc_family="L1"
2828..3176
/rpc_family="MaLR"
3217..3544
/rpc_family="Alu"
3617..3648
/rpc_family="AT_rich"
3793..4078
/rpc_family="Alu"
4138..4267
/rpc_family="Alu"
4462..4655
/rpc_family="MER1_type"
4639..4688
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"

```

## Alignment Scores:

```

Pred. No.: 1,43e-26 Length: 184206
Score: 471.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.66% Indels: 0
DB: 9 Gaps: 0

```

US-09-809-638-2 (1-699) x AC020593 (1-184206)

```

QY 267 ThrGlyThrAlaSerAlaIacIleuLeuYrYLeuHisThrTrpAlaAlaValSer 286
|||||
DB 111216 ACAGGACAGCTTCAGCTGCGGGGCTCTTACCTGCACACATGGGACAGCTCTGTCT 111275

```





```

misc_feature 27006..29581
/note="assembly_fragment:00175"
misc_feature 29682..33446
/note="assembly_fragment:00178"
misc_feature 33547..48536
/note="assembly_fragment:00416"
misc_feature 48637..55359
/note="assembly_fragment:00542"
misc_feature 55700..65124
/note="assembly_fragment:00565"
misc_feature 65225..69090
/note="assembly_fragment:00583"
misc_feature 69191..73894
/note="assembly_fragment:00649"
misc_feature 73995..82665
/note="assembly_fragment:00709"
misc_feature 82766..87239
/note="assembly_fragment:00789"
misc_feature 87340..93671
/note="assembly_fragment:00874"
misc_feature 93772..97550
/note="assembly_fragment:00924"
misc_feature 97651..100101
/note="assembly_fragment:01052"
misc_feature 100202..103416
/note="assembly_fragment:01082"
misc_feature 103517..107564
/note="assembly_fragment:01216"
misc_feature 107665..131778
/note="assembly_fragment:01234"
misc_feature 131879..135261
/note="assembly_fragment:01280"
misc_feature 135362..137913
/note="assembly_fragment:01337"
misc_feature 138014..140949
/note="assembly_fragment:01643"
misc_feature 141050..156466
/note="assembly_fragment:01660"
misc_feature 156567..164836
/note="assembly_fragment:01690"
misc_feature 164937..168950
/note="assembly_fragment:01692.0"
misc_feature 169051..173724
/note="assembly_fragment:01748"
misc_feature 173825..179991
/note="assembly_fragment:02061"
misc_feature 180092..190208
/note="assembly_fragment:02140"
misc_feature 190309..193471
/note="assembly_fragment:02150"
misc_feature 193572..197829
/note="assembly_fragment:02181"
misc_feature 197930..222885
/note="assembly_fragment:00834
clone_end:77
vector_side:right"

```

BASE COUNT 69666 a 41364 c 44947 g 63682 t 3226 others  
ORIGIN

Alignment Scores: 1.97e-11 Length: 222885  
Pred. No.: 287.50 Matches: 63  
Score: 70.30% Conservative: 8  
Percent Similarity: 62.38% Mismatches: 15  
Best Local Similarity: 7.73% Indels: 15  
Query Match: 2 Gaps: 3

US-09-809-638-2 (1-699) x AL158816 (1-222885)

Oy 606 ThrAspHisAspArgTrpCysGluTrpLeuMetTyrArgGlyLeuIleArgLeuGlyTyr 625  
|||||  
Db 157651 ACAAAACATTTCTGTTCTGC-----TACAGGTTGGCTTAT 157686

```

Oy 626 AAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMetAlaLysPheArgIle 645
|||||
Db 157687 GCAAGAACTCCCACTGACCTGAGTATTCAGAAATTCAGATGCAAAATTTAGAGATC 157746
|||||
Oy 646 ProAspAspProThrAsnTrpArgAspAsnGlnLysValValIleAspHisArgGlyVal 665
|||||
Db 157747 CCTGATGACCCACCTAAATTAATAGACACACAGAAAGTGTCATGACACAGAGAAGTT 157806
|||||
Oy 666 SerGluLysIleHisPheAsnProArg-----PheGlySerTyrLys 679
|||||
Db 157807 TCTGCAAAATTCATTTTATTCACCAAGTACGTTCTTATGCTGAGTCCAGTTATGAG 157866
|||||
Oy 680 GluGlyHis---AsnTrpGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPhe 698
|||||
Db 157867 CTACTGCATCCCTCATGAGACCTACATATTAACCTCAAAATGAGCTAGAAATTAATTTT 157926
|||||
Oy 699 Leu 699
Db 157927 GTC 157929

```

```

RESULT 9
LOCUS AL158816/c
DEFINITION Homo sapiens chromosome 6 clone RP3-405B4, ** SEQUENCING IN
ACCESSION AL158816
VERSION AL158816.11 GI:9943990
KEYWORDS HTG: HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Mammali; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE 1 (bases 1 to 222885)
JOURNAL

```

COMMENT Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Aug 29, 2000 this sequence version replaced gi:9926474.

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

Project Information

Center project name: dd405B4

Summary Statistics

Sequencing program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 204390 bases at least Q40

Consensus quality: 211527 bases at least Q40

Insert size: 219685; sum-of-contigs

Insert size: 126545; 14.0% error; agarose-fp

Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality

coverage: 6.72x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently

consists of 33 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 3599: contig of 3599 bp in length

3600 3699: gap of 100 bp

3700 12818: contig of 9119 bp in length

12819 12918: gap of 100 bp

12919 16398: contig of 3480 bp in length

16399 16498: gap of 100 bp

16499 18590: contig of 2092 bp in length

```
* 18591 18690: gap of 100 bp
* 18691 22372: contig of 3682 bp in length
* 22373 22472: gap of 100 bp
* 22473 26905: contig of 4433 bp in length
* 26906 27005: gap of 100 bp
* 27006 29581: contig of 2576 bp in length
* 29582 29681: gap of 100 bp
* 29682 33446: contig of 3765 bp in length
* 33447 33546: gap of 100 bp
* 33547 48536: contig of 14990 bp in length
* 48537 48636: gap of 100 bp
* 48637 55599: contig of 6963 bp in length
* 55600 55699: gap of 100 bp
* 55700 65124: contig of 9425 bp in length
* 65125 65224: gap of 100 bp
* 65225 69090: contig of 3866 bp in length
* 69091 69190: gap of 100 bp
* 69191 73894: contig of 4704 bp in length
* 73895 73994: gap of 100 bp
* 73995 82665: contig of 8671 bp in length
* 82666 82765: gap of 100 bp
* 82766 87239: contig of 4474 bp in length
* 87240 87339: gap of 100 bp
* 87340 93671: contig of 6332 bp in length
* 93672 93771: gap of 100 bp
* 93772 97550: contig of 3779 bp in length
* 97551 97650: gap of 100 bp
* 97651 100101: contig of 2451 bp in length
* 100102 100201: gap of 100 bp
* 100202 103416: contig of 3215 bp in length
* 103417 103516: gap of 100 bp
* 103517 107564: contig of 4048 bp in length
* 107565 107664: gap of 100 bp
* 107665 131778: contig of 2414 bp in length
* 131779 131878: gap of 100 bp
* 131879 135261: contig of 3383 bp in length
* 135262 135361: gap of 100 bp
* 135362 137913: contig of 2552 bp in length
* 137914 138013: gap of 100 bp
* 138014 140949: contig of 2936 bp in length
* 140950 141049: gap of 100 bp
* 141050 156466: contig of 15417 bp in length
* 156467 156566: gap of 100 bp
* 156567 164836: contig of 8270 bp in length
* 164837 164936: gap of 100 bp
* 164937 168950: contig of 4014 bp in length
* 168951 169050: gap of 100 bp
* 169051 173724: contig of 4674 bp in length
* 173725 173824: gap of 100 bp
* 173825 179991: contig of 6167 bp in length
* 179992 180091: gap of 100 bp
* 180092 190208: contig of 10117 bp in length
* 190209 190308: gap of 100 bp
* 190309 193471: contig of 3163 bp in length
* 193472 193571: gap of 100 bp
* 193572 197829: contig of 4258 bp in length
* 197830 197928: gap of 100 bp
* 197929 222885: contig of 24956 bp in length.
* 197930
```

## FEATURES

## Source

```
1. 222885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="RP3-405B4"
/clone_lib="RPC1-3"
1. 3599
/misc_feature
/note="assembly_fragment:00027"
3700. 12818
/misc_feature
/note="assembly_fragment:00056"
12919. 16398
/misc_feature
/note="assembly_fragment:00093"
16499. 18590
/misc_feature
/note="assembly_fragment:00125"
```

```
misc_feature 18691. 22372
/note="assembly_fragment:00140"
misc_feature 22473. 26905
/note="assembly_fragment:00167"
misc_feature 27006. 29581
/note="assembly_fragment:00175"
misc_feature 29682. 33446
/note="assembly_fragment:00178"
misc_feature 33547. 48536
/note="assembly_fragment:00416"
misc_feature 48637. 55599
/note="assembly_fragment:00542"
misc_feature 55700. 65124
/note="assembly_fragment:00565"
misc_feature 65225. 69090
/note="assembly_fragment:00583"
misc_feature 69191. 73894
/note="assembly_fragment:00649"
misc_feature 73995. 82665
/note="assembly_fragment:00709"
misc_feature 82766. 87239
/note="assembly_fragment:00789"
misc_feature 87340. 93671
/note="assembly_fragment:00874"
misc_feature 93772. 97550
/note="assembly_fragment:00924"
misc_feature 97651. 100101
/note="assembly_fragment:01052"
misc_feature 100202. 103416
/note="assembly_fragment:01082"
misc_feature 103517. 107564
/note="assembly_fragment:01216"
misc_feature 107665. 131778
/note="assembly_fragment:01234"
misc_feature 131879. 135261
/note="assembly_fragment:01280"
misc_feature 135362. 137913
/note="assembly_fragment:01337"
misc_feature 138014. 140949
/note="assembly_fragment:01643"
misc_feature 141050. 156466
/note="assembly_fragment:01660"
misc_feature 156567. 164836
/note="assembly_fragment:01690"
misc_feature 164937. 168950
/note="assembly_fragment:01692.0"
misc_feature 169051. 173724
/note="assembly_fragment:01748"
misc_feature 173825. 179991
/note="assembly_fragment:02061"
misc_feature 180092. 190208
/note="assembly_fragment:02140"
misc_feature 190309. 193471
/note="assembly_fragment:02150"
misc_feature 193572. 197829
/note="assembly_fragment:02181"
misc_feature 197930. 222885
/note="assembly_fragment:00834"
misc_feature 197930
vector_end:77
vector_side:right"
BASE COUNT 69666 a 41364 c 44947 g 63682 t 3226 others
ORIGIN
```

## Alignment Scores:

```
Pred. No.: 7.09e-10 Length: 222885
Score: 268.50 Matches: 58
Percent Similarity: 70.93% Conservative: 3
Best Local Similarity: 67.44% Mismatches: 16
Query Match: 7.22% Indels: 9
DB: 2 Gaps: 2
```

US-09-809-638-2 (1-699) x AL158816 (1-222885)

Oy	457	ThrGluAlaAspPheIleThrIleuGluSerAspAlaSerIleProIleMetGlyAsn	476
Db	217412	ACAGGTGCAGATTTCATACAAATTTTGGAGATGATGCTTCACAGCCCTATATGGGAAAC	217413
Oy	477	AsnAspLeuThrMetIleProLeuGlyGluIuIysLeuGlyPheTyThrAspPheGlyProSer	496
Db	217412	AATGCTTAACCATGTGGCTAGGGGAAAAAGTTGGGTTCCTATACAGACTTTGGTCCAAAC	217353
Oy	497	ThrArgTyHisIleThrTropolyIleMetAlaLeu-----SerArgTyIleProIleVal	513
Db	217352	ACAAAGGTATCACACTGGGGGGGAGACTATACCTTTGGGAGTTAAATCCGAGAGATAGTCATGA	217293
Oy	514	LysSerGluHisIleLeuLeuProSerProGlu-----GlyGluIle	527
Db	217292	AAGGCTAGGCTGCATAGGCTTGAGATGGCAAACTTGATTTGTAATATCTTTGGTGAATGT	217233
Oy	528	AlaProAlaIleThrLeu 533	
Db	217232	GGCCCAAGTGATGCTTCCTT 217215	

[illegible]

**ACCESSION** AL512496  
**VERSION** AL512496.1  
**KEYWORDS** 1 norganic phosphate transporter; metaxin 2; mitochondrial outer

SOURCE membrane protein; phosphatidylinositol metabolism; spindle integrity WD domain; ubiquitin-like protein.  
fission yeast.

**ORGANISM** Schizosaccharomyces pombe  
Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;

REFERENCE  
AUTHORS

Schizosaccharomyces.  
1 (bases 1 to 18342)  
Wood, V., Rajandream, M. A., Barrell, B. G., Aert, R., Robben, J.,

**TITLE**  
**JOURNAL**  
**Submitted (01-JAN-2001)** European Schizosaccharomyces genome  
Welljens, I., Grymonprez, B. and Volckaert, G.  
Direct Submission

sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barreillesanger.ac.uk and Katholieke Universiteit Leuven, Laboratory of Gene Technology,

COMMENT

Kardinaal Mercierlaan, 9  
Notes:  
Details of yeast sequencing at the Sanger Centre are available on

the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/))  
During 1995 to 1996 about 66% of *S. pombe* chromosome I was

sequenced by the Sanger Centre. The sequencing of the *S. pombe* genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including

the Sanger Centre, are participating in the project. Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database).

database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the

number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC25H2.01c, SP (S. pombe), B (chromosome 2), C25H2

(cosmid name), .01 (first CDS), .c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length

in codons is given for each CDS.  
**IMPORTANT:** This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence

overlapping sections may be better, because we arrange for a small overlap between neighbouring submissions. Cosmid 589 is overlapped at the 5' end by cosmid 16A10. FMT entry STAC16A10

FEATURES  
Location/Qualifiers  
accession number 297185, and at the 3' end by cosmid c688, EMBL entry SPAC688, accession number AL355632.

source . location/variables  
1. .18342

```

/organism="Schizosaccharomyces pombe"
/strain="972h"
/db_xref="taxon:4896"
/chromosome="I"
/map="IL"
/locus="Drocm1d.0500"

```

```

/cellno= cosmid c369
1. .89
/note="nominal overlap with SPAC16a10 S. pombe chromosome
1"

```

```
complement(join(1, .65, 121, .777))
/gene="SPAC589.01c"
/ncbi="SPAC16210.00c"
```

```

/note-3fac16a10.06c
complement(join(1. .65,121. .777))
/partial
/cross="SPM0590_01a"

```

/gene=SPAC589.01c, very low SIMILARITY:Caenorhabditis elegans., Q21148, k02f3.4 protein., (308 aa), fasta format. contig111 EV1234 (3238) (positive) (in 102 aa)

```

scores: opt: III, E():2.4, (23.38 identity in 155 aa)
/codon_start=1
/label=SPAC589.01c
/product="Phenothiazine"

```

```

/product=hypothetical protein
/protein_id="CAC19758.1"
/db_xref="GI:12043545"

```

SRLDSSHNSGLDALAMTPYIVKENNKTHTLSPRGNYPSAEKSCDFYQGPLHYRENTS

PCLYGSSLSNVTEKSSDYKHGLGSPISLVSSNIASKLHGRKDIYKLSLELRNCINSS  
KDSNQIQELENLRLRKEKQRNTE"  
complement(66,.81)  
12345678910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210310410510610710810911011111211311411511611711811912012112212312412512612712812913013113213313413513613713813914014114214314414514614714814915015115215315415515615715815916016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221321421521621721821922022122222322422522622722822923023123223323423523623723823924024124224324424524624724824925025125225325425525625725825926026126226326426526626726826927027127227327427527627727827928028128228328428528628728828929029129229329429529629729829930030130230330430530630730830931031131231331431531631731831932032132232332432532632732832933033133233333433533633733833934034134234334434534634734834935035135235335435535635735835936036136236336436536636736836937037137237337437537637737837938038138238338438538638738838939039139239339439539639739839940040140240340440540640740840941041141241341441541641741841942042142242342442542642742842943043143243343443543643743843944044144244344444544644744844945045145245345445545645745845946046146246346446546646746846947047147247347447547647747847948048148248348448548648748848949049149249349449549649749849950050150250350450550650750850951051151251351451551651751851952052152252352452552652752852953053153253353453553653753853954054154254354454554654754854955055155255355455555655755855956056156256356456556656756856957057157257357457557657757857958058158258358458558658758858959059159259359459559659759859960060160260360460560660760860961061161261361461561661761861962062162262362462562662762862963063163263363463563663763863964064164264364464564664764864965065165265365465565665765865966066166266366466566666766866967067167267367467567667767867968068168268368468568668768868969069169269369469569669769869970070170270370470570670770870971071171271371471571671771871972072172272372472572672772872973073173273373473573673773873974074174274374474574674774874975075175275375475575675775875976076176276376476576676776876977077177277377477577677777877978078178278378478578678778878979079179279379479579679779879980080180280380480580680780880981081181281381481581681781881982082182282382482582682782882983083183283383483583683783883984084184284384484584684784884985085185285385485585685785885986086186286386486586686786886987087187287387487587687787887988088188288388488588688788888989089189289389489589689789889990090190290390490590690790890991091191291391491591691791891992092192292392492592692792892993093193293393493593693793893994094194294394494594694794894995095195295395495595695795895996096196296396496596696796896997097197297397497597697797897998098198298398498598698798898999099199299399499599699799899910001001100210031004100510061007100810091010101110121013101410151016101710181019102010211022102310241025102610271028102910301031103210331034103510361037103810391040104110421043104410451046104710481049105010511052105310541055105610571058105910601061106210631064106510661067106810691070107110721073107410751076107710781079108010811082108310841085108610871088108910901091109210931094109510961097109810991100110111021103110411051106110711081109111011111112111311141115111611171118111911201121112211231124112511261127112811291130113111321133113411351136113711381139114011411142114311441145114611471148114911501151115211531154115511561157115811591160116111621163116411651166116711681169117011711172117311741175117611771178117911801181118211831184118511861187118811891190119111921193119411951196119711981199120012011202120312041205120612071208120912101211121212131214121512161217121812191220122112221223122412251226122712281229123012311232123312341235123612371238123912401241124212431244124512461247124812491250125112521253125412551256125712581259126012611262126312641265126612671268126912701271127212731274127512761277127812791280128112821283128412851

```

/gene="SPAC589.01c"
/note="ctgaccaagcgaatag, splice branch and acceptor"
complement(115..120)

```

```

/ gene="SPAC589.01C"
/ note="gtacgt, splice donor sequence"
complement(join(967, .4482,4518, .4667,4770, .4775))

```

```

/ gene="SPAC589.02c"
complement(join(967.
/ gene="SPAC589.02c"

```

```
/note="SPAC589.02c, len:1224"  
/codon_start=1  
/label=SPAC589.02c
```

```

/product=hypothetical protein"
/protein_id="CAC19759.1"
/db_xref="GI:12043546"

```

```

/translation=-MSVCFKRYKRWPHVLSRYASIFLTNNLPRESSAMHDG
HLISIKTILTDOPCNTYYAKYQVRKDNEEDASLLEDKAYLRNDCIVAHKNDLL
FVYDFQAIPIPESSSEFLLNSGAFSRLALFQKDEALLDLDLTINFLQGLKKTLYW

```

VGVTLESSALEFSPLELCFLETSTALRMDSDLSNLJDLNMLVLPSPNYNNKKELTE

FANEAEASDKRGEFTEKEETADAVTLVPSHSSPVNSINSAKSTPASIKVNEEI  
LVADHNVSDDI LMEEIDDVGITTEADFDYFDL.PNVEEKVEMIEPNFANTMTTLDNEEIN  
TISQSNTSPNLNTHENIPKQMEIQSDDRMYTEDLNPYNVEVDIPELSJENISDSKIPT

SAYMPSYYSAVIFPSSISSIFQKYNVGKWCPSPLSTEDLEESFSAESTSTDED  
ICSTNFIQQDFTMEYNHDEFSSSKPTNISSEQSNPDSNYDTLSLAHQVLMNESKANF  
DSFELKSLDLOPTITLTKNDLLNAILSQNLWFRSLPFWKSMTTTFMMSQDVLFNFSSYM

RKPIRDYLENILLGESSAVFLSKSPENYLSSINNGHALNDNPSSQVNFSETLVNFQ  
PPRVLTKYNEKRLSDSSAPENWISLCLQPYGESKDFEYFLSSKSPDYSSKAISSFY  
DVQLAYENCKLGKLNLSSETSINERVMGFSTINETDNYDNETTQSDPATSYEQLASV

CVNELSGKNLEFFYFLEDDSEKLLKACQHFICVKDSIKRLGDNKFEDKSLRICTIPNS  
IFDSPNSHTTNSNSFFTKVSLDIYNNDFPLMDGSLKREPAFLKKPLSTLNYQLKD  
INPRSSALGEYALHVYTTYVEEHLICWNDSYGEFETERRYFLODEIEDALOQILE

VTFETFLNSMMDWIVIMKIGEMSDAEYLFWDQAIIPENLOGNVSLTVGYCSAEHGP  
STSKVFESRIPYASASTVIRNNSHLSVAFIREMAMPVNDFFKISTITLARGYAL  
DEDESYLPPLSIHLISRNDPYLMLNLIKHYLSMITYLQRTVVSFSSLPHISTVL

```
YQKQLQFMASDITHPVTS"
complement(4483..4502)
/gene="SPAC589.02c"
```

```

/note="atacatgataaattaaaag, splice branch and acceptor"
complement(4512..4517)
/gene="SPAC589.02c"

```

```

/note="gtatt, splice donor sequence"
complement(4668..4686)

```



QY 293 ethr-----AlasermetrpProGlnThrLeuGlyHisLeuIleasnse 308  
 Db 17980 CGTTTAATTACTCGTGGCTTCATC-----AGATTTCCTC 18018  
 QY 308 rglYthraPnProGlyLysThrmethrIleAlaMetIlePheTyLeuLeuGluIleph 328  
 Db 18019 CTTCATTCCTCCGCGAGATTGGGCGGCTTTTTCGTTTTCATCATCTACTACT 18078  
 QY 328 ePheCySaLatPcYsThAlaPheLysPheValProGlyGlyValTyAlaArgLuar 348  
 Db 18079 AGCTCATGTTTGGGTGCTGACCTCAGCATGTTGTCAGGTGAGCCATTCTTCGAGAGAG 18138  
 QY 348 gSerAspValLeuGlyThrmethrLeuIleGlyLeuAsnmetLeuPhe----- 366  
 Db 18139 AACTTCTTACATTCATC-----ATTTCATTTGGTGGACCTTCGACGCTTAGT 18186  
 QY 367 -----GlyProLysLysAsnLeuAspLeuLeuGluIleTh 378  
 Db 18187 ACCTGCGATTCGGGTGAGTCCAAAGACCAAGACGACTCATCTAGTATGATAT 18246  
 QY 376 rLysAsnSerLysValLeuPheArgLysSer-GluLys-----TyrMetLysLeuP 396  
 Db 18247 CAACACATCAGATCTTCTTATTCGTAGAGAGAGTTTCAAAAAATCATATTCACAGCTTT 18306  
 QY 396 helLeuTrpLeuLeu 400  
 Db 18307 TTGTCTGGCTCTTA 18320

RESULT 11  
 AC106576/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-198N10, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 68 unordered pieces.  
 AC106576 165893 bp DNA linear HTG 12-JAN-2002  
 AC106576.1 GI:18139100  
 HTG: HTGS\_PHASE1.  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 165893)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,  
 Butcher,P., Burkelt,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Deen,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan,Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
 Hollins,B., Homsl,F., Howard,S., Huber,D., Huix,S., Hume,J.,  
 Johnson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,D.,  
 Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,  
 Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Peedy,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,

COMMENT

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,  
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Swalek,A., Taber,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Washington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 165893)  
 Worley,K.C.  
 Direct Submission  
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GLJY  
 Center clone name: CH230-198N10  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329First call to  
 findPhrap.plst  
 Consensus quality: 137306 bases at least Q40  
 Consensus quality: 144043 bases at least Q30  
 Consensus quality: 149677 bases at least Q20  
 Estimated insert size: 139846; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-gel estimation  
 Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 68 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1  
 \* 7957 7956: contig of 7956 bp in length  
 \* 8057 8056: gap of unknown length  
 \* 13904 13903: contig of 5847 bp in length  
 \* 13904 14003: gap of unknown length  
 \* 14004 17958: contig of 3955 bp in length  
 \* 17959 18058: gap of unknown length  
 \* 18059 22354: contig of 4296 bp in length  
 \* 22355 22454: gap of unknown length  
 \* 22455 26589: contig of 4135 bp in length  
 \* 26590 31248: gap of unknown length  
 \* 31249 31348: contig of 4559 bp in length  
 \* 31349 35556: gap of unknown length  
 \* 35557 36556: contig of 5208 bp in length  
 \* 36557 40383: gap of unknown length  
 \* 40384 44415: contig of 3727 bp in length  
 \* 44416 44515: gap of unknown length  
 \* 44516 49038: contig of 4523 bp in length  
 \* 49039 52349: gap of unknown length  
 \* 52350 56176: gap of unknown length  
 \* 56177 60332: contig of 3727 bp in length  
 \* 60333 60432: gap of unknown length  
 \* 60433 62803: contig of 4056 bp in length  
 \* 62804 62903: gap of unknown length  
 \* 62903: gap of unknown length

62904	66621: contig of 3718 bp in length	142433	142532: gap of unknown length	66622	66721: gap of unknown length	142533	144347: contig of 1815 bp in length	66722	69020: contig of 2299 bp in length	144348	144447: gap of unknown length	69021	69120: gap of unknown length	144449	145945: contig of 1498 bp in length	69121	72220: contig of 3100 bp in length	145946	146045: gap of unknown length	72221	72321: gap of unknown length	146046	147800: contig of 1755 bp in length	72322	75607: contig of 3287 bp in length	147801	147900: gap of unknown length	75608	75707: gap of unknown length	147901	149217: contig of 1317 bp in length	75709	78960: contig of 3253 bp in length	149218	149317: gap of unknown length	78961	81780: contig of 2720 bp in length	149318	150337: contig of 1020 bp in length	81781	81880: gap of unknown length	150338	150437: gap of unknown length	81881	85140: contig of 3260 bp in length	150438	151919: contig of 1482 bp in length	85141	85240: gap of unknown length	151920	152019: contig of 1482 bp in length	85241	87883: contig of 2643 bp in length	152020	152119: contig of 1482 bp in length	87884	87984: contig of 2779 bp in length	152120	152219: contig of 1482 bp in length	87985	90762: contig of 2779 bp in length	152221	152319: contig of 1482 bp in length	90763	90862: gap of unknown length	152320	152419: contig of 1482 bp in length	90863	93086: contig of 2224 bp in length	152420	152519: contig of 1482 bp in length	93087	93186: gap of unknown length	152520	152619: contig of 1482 bp in length	93187	96030: contig of 2844 bp in length	152621	152719: contig of 1482 bp in length	96031	96130: gap of unknown length	152720	152819: contig of 1482 bp in length	96131	98070: contig of 1940 bp in length	152821	152919: contig of 1482 bp in length	98071	98170: gap of unknown length	152920	153019: contig of 1482 bp in length	98171	100910: contig of 2740 bp in length	153020	153119: contig of 1482 bp in length	100911	101010: gap of unknown length	153120	153219: contig of 1482 bp in length	101011	102968: contig of 1958 bp in length	153220	153319: contig of 1482 bp in length	102969	103068: gap of unknown length	153320	153419: contig of 1482 bp in length	103069	104748: contig of 1680 bp in length	153420	153519: contig of 1482 bp in length	104749	104848: gap of unknown length	153520	153619: contig of 1482 bp in length	104849	106306: contig of 1458 bp in length	153621	153719: contig of 1482 bp in length	106307	106406: gap of unknown length	153720	153819: contig of 1482 bp in length	106407	108402: contig of 1996 bp in length	153821	153919: contig of 1482 bp in length	108403	108502: gap of unknown length	153920	154019: contig of 1482 bp in length	108503	110334: contig of 1732 bp in length	154020	154119: contig of 1482 bp in length	110335	110334: gap of unknown length	154120	154219: contig of 1482 bp in length	110335	112075: contig of 1741 bp in length	154221	154319: contig of 1482 bp in length	112076	112175: gap of unknown length	154320	154419: contig of 1482 bp in length	112176	113608: contig of 1433 bp in length	154421	154519: contig of 1482 bp in length	113609	113708: gap of unknown length	154520	154619: contig of 1482 bp in length	113709	115299: contig of 1591 bp in length	154621	154719: contig of 1482 bp in length	115300	115399: gap of unknown length	154720	154819: contig of 1482 bp in length	115301	116700: contig of 1271 bp in length	154821	154919: contig of 1482 bp in length	116701	116770: gap of unknown length	154920	155019: contig of 1482 bp in length	116771	119076: contig of 2306 bp in length	155020	155119: contig of 1482 bp in length	119077	119176: gap of unknown length	155120	155219: contig of 1482 bp in length	119177	120752: contig of 1576 bp in length	155221	155319: contig of 1482 bp in length	120753	120852: gap of unknown length	155320	155419: contig of 1482 bp in length	120853	122336: contig of 1484 bp in length	155421	155519: contig of 1482 bp in length	122337	122436: gap of unknown length	155520	155619: contig of 1482 bp in length	122437	125169: contig of 2733 bp in length	155621	155719: contig of 1482 bp in length	125170	125269: gap of unknown length	155720	155819: contig of 1482 bp in length	125270	126815: contig of 1546 bp in length	155821	155919: contig of 1482 bp in length	126816	126915: gap of unknown length	155920	156019: contig of 1482 bp in length	126916	128187: contig of 1272 bp in length	156021	156119: contig of 1482 bp in length	128188	128287: gap of unknown length	156120	156219: contig of 1482 bp in length	128288	129839: contig of 1552 bp in length	156221	156319: contig of 1482 bp in length	129839	129939: gap of unknown length	156320	156419: contig of 1482 bp in length	129940	131660: contig of 1721 bp in length	156421	156519: contig of 1482 bp in length	131661	131760: gap of unknown length	156520	156619: contig of 1482 bp in length	131761	133894: contig of 2134 bp in length	156621	156719: contig of 1482 bp in length	133895	133994: gap of unknown length	156720	156819: contig of 1482 bp in length	133995	135658: contig of 1664 bp in length	156821	156919: contig of 1482 bp in length	135659	135758: gap of unknown length	156920	157019: contig of 1482 bp in length	135759	136662: contig of 1104 bp in length	157021	157119: contig of 1482 bp in length	136663	136962: gap of unknown length	15712
-------	------------------------------------	--------	-------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------------	-------	------------------------------------	--------	-------------------------------------	-------	------------------------------	--------	-------------------------------------	-------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------------	--------	-------------------------------	-------

QY 684 TYRGLAASNAHSHISHISHMETASNTHPROLYSTYR 697  
 DB 240 TATGAAAACACCATCATTTTCATATGATGATCCCAATATAC 199

RESULT 13  
 CNS06EJ2

LOCUS CNS06EJ2 849 bp DNA linear STS 10-JAN-2001  
 DEFINITION T3 end of clone AROA022A01 of library AROA from strain CBS 732 of  
 Zygosaccharomyces rouxii, sequence tagged site.

ACCESSION AL395172  
 VERSION AL395172.1 GI:12146608  
 KEYWORDS STS.  
 SOURCE Zygosaccharomyces rouxii.  
 ORGANISM Zygosaccharomyces rouxii.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.  
 1 (bases 1 to 849)  
 de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,  
 Wincker,P., Artiguenave,F. and Souciet,J.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 8.  
 Zygosaccharomyces rouxii(1)  
 FEBS Lett. 487 (1), 52-55 (2000)

REFERENCE  
 JOURNAL 11132863  
 PUBLISHED 2 (bases 1 to 849)  
 REFERENCE Souciet,J.L., Artiguenave,F., Blandin,G.,  
 Auteurs Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
 de Montigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of  
 yeast species for molecular evolution studies(1)  
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE  
 JOURNAL 11152876  
 PUBLISHED 3 (bases 1 to 849)  
 REFERENCE Genoscope.  
 Auteurs Direct Submission  
 JOURNAL Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 sequefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomycetes  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 anomala, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

COMMENT  
 Location/Qualifiers  
 1. 849  
 /organism="Zygosaccharomyces rouxii"  
 /strain="CBS 732"  
 /db\_xref="taxon:4956"  
 /clone="AROAO22A01"  
 /clone.lib="AROAA"  
 /note="end : T3"  
 misc\_feature  
 <5..>847  
 /note="similar to Saccharomyces cerevisiae ORF YCR017c [ similarity to hypothetical S.pombe protein ]"  
 /evidence=not-experimental

BASE COUNT 198 a 157 c 193 g 300 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0014 Length: 849  
 Score: 152.50 Matches: 71  
 Percent Similarity: 33.23% Conservative: 34  
 Best Local Similarity: 22.47% Mismatches: 120  
 Query Match: 4.10% Indels: 91  
 DB: 11 Gaps: 13

US-09-809-638-2 (1-699) x CNS06EJ2 (1-849)

QY 87 ProAnaAlaValLeuArgLeuMetValLeuAlaLeuGlyValSerSerLeuLeuVal 106  
 DB 17 CCGGAAGTGGCTG-----CTGGACTGGTCCGGATGCTCTTGTGGTG 64

QY 107 GlnAla----- 108  
 DB 65 ACTGCATTGTGTCATCAATTTGAATCTATTAAAGACCTGAAGTCATGCGCCTTCACT 124

QY 109 ValThrTPTPSPSerGlySerHisLeuGlnArgTyrLeuArgIleTPPcIlyPheIleLeu 128  
 DB 125 CTGACATG-----TTGTTG 139

QY 129 GlyGlnIleValLeuValValLeuArgIleTPPcIlyPheIleLeu 147  
 DB 140 GGTGTGTCATCTCTGTGATATTAAATTCGCTGCCATGTAACAAATCCCTTGCGCT 199

QY 148 -----SerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160  
 DB 200 ATTACTAAAGCTGAAGATGCTGCTATGCACACATTTGATTTTCACAAACATTTT 259

QY 161 AlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCysSerLysProGluGly 180  
 DB 260 GGAATCTTGCA-----CCTATGTCAATTCATATGATTTTGCCTCCAAAGATAGTTCTAAG 316

QY 181 LysThrGlyGlyValAlaThrGlyMetAlaSerArgProAsnTPPLeuAlaGlyAla 200  
 DB 317 GCTGACGGAACACTACGGAATCTTAAAGAGTAACAATACAACTTTTGTGGGCTCT 376

QY 201 AlaPheGlySerLeuValPheLeuThrHisTPPValPheGlyGlyValSerLeuValSer 220  
 DB 377 GGATTCGGTTCATATATTTTGTCTTCACGATTCCTTACCGATGCGCTGCATGATT 436

QY 221 ArgTPPAlaValSerGly-----HisProHisProGlyProAsnProPheGly 238  
 DB 437 TATGTGCTCTGGAGGCTTGGAGAACCAAGAACTCAAGGCCCTTGCTTGGCATGGAGT 496

QY 239 GlyAlaValLeuLeuCysLeuAlaSerGlyLeu-----MetLeuProSerCysLeuTPPhe 257  
 DB 497 GGG-----TTACTTTCGCTTGTCTCATGCTTATAGCTTCTTAACTCCGTGAGATTCCT 550

QY 258 ArgGlyThrGlyLeuIleTPPThrPValThrGlyThrAlaSerAlaIleGlyLeuLeuTyr 277  
 DB 551 AAAAATACA-----TGCTACTCGGTATC-----CTTTGTTG 583

QY 278 LeuHisThrTPPAlaAla----- 284  
 DB 584 GAATCTACTTGGGCTTGTGCAATGAAACATTAATGATGAGATTAATCTTTACGCT 643

QY 285 -----ValSerGlyCysValPheAlaIle-----PheThrAlaSerMet 297  
 DB 644 GGGTACCTTATATCTTGGGTGTGATTTGGACCATTTTCATATTTTGGACCTTTGAGT 703

QY 298 TrpProGlnThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThr 317  
 DB 704 GAATTACAGAGGTGGCGGTATAT----- 730

QY 318 IleAlaMetIlePheTyrLeuLeuGluIlePhePheCysAlaTPPcIlyPheAlaPheLys 337  
 DB 721 CTAGCTTTTGTACCTATGCGCTTCATGCTTATTTGGCAGCTGTTGCGATGCA 790

QY 338 PheValProGlyGlyValTyrAlaArgGluArgSerAspValLeuLeu 353  
 DB 791 TTTGTCTCTTGATGTGTTATGAGAGAACATGAGAGCTGTATTG 838

RESULT 14  
 AC096364/c AC096364 22532 bp DNA linear HTG 20-DEC-2001  
 LOCUS Rattus norvegicus chromosome R1 clone CH230-59C15, WORKING DRAFT  
 DEFINITION  
 SEQUENCE, 20 unordered pieces.  
 ACCESSION AC096364



VERSION AC096364.4 GI:17944066  
 KEYWORDS HTGS: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLLTOP.  
 SOURCE Norway rat.  
 ORGANISM *Rattus norvegicus*  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 225532)  
 Munz, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
 Alshrocks, S.L., Amaratunga, H.C., Are, J.R., Banks, F., Barbata, J.,  
 Benton, J., Bimaga, K., Blankenbury, K., Bonnin, D., Bouck, J.,  
 Bowie, S., Brieva, M., Brown, E., Brown, N.P., Buhay, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carrott, T.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R.,  
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
 Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, Y., Dinn, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Dublin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunatirane, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,  
 Hollins, B., Homai, F., Howard, S., Huber, J., Huiyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J.,  
 Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W.,  
 Louisse, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, J., Luna, R.,  
 Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S.,  
 Ogih, M., Okunou, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, M., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,  
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitai, N.,  
 Slisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gdbbs, R.  
 Unpublished  
 Direct Submission  
 2 (bases 1 to 225532)  
 Worley, K.C.  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:15920027.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GEZO  
 Center clone name: CH230-59C15  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329first call to  
 findPhraplist  
 Consensus quality: 220440 bases at least Q40  
 Consensus quality: 221708 bases at least Q30  
 Consensus quality: 222752 bases at least Q20  
 Estimated insert size: 216271; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 10.2x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 20 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1	37343	contig of 37343 bp in length
*	37443	gap of unknown length
*	37444	59505: contig of 22062 bp in length
*	59506	59506: gap of unknown length
*	59507	59507: gap of unknown length
*	59508	59508: contig of 37303 bp in length
*	96909	96909: gap of unknown length
*	97009	126921: contig of 29913 bp in length
*	126922	127021: gap of unknown length
*	127022	144976: contig of 17955 bp in length
*	144977	145076: gap of unknown length
*	145077	145076: gap of unknown length
*	145078	145076: contig of 18872 bp in length
*	163949	164048: gap of unknown length
*	163949	164048: gap of unknown length
*	164049	170970: contig of 6922 bp in length
*	170971	171070: gap of unknown length
*	171071	179611: contig of 8541 bp in length
*	179612	179711: gap of unknown length
*	179712	186544: contig of 6833 bp in length
*	186545	186644: gap of unknown length
*	186645	192903: contig of 6259 bp in length
*	192904	193003: gap of unknown length
*	193004	201690: contig of 8687 bp in length
*	201691	201790: gap of unknown length
*	201791	205903: contig of 5113 bp in length
*	205904	207003: gap of unknown length
*	207004	209367: contig of 2364 bp in length
*	209368	209467: gap of unknown length
*	209468	214422: contig of 4955 bp in length
*	214423	214522: gap of unknown length
*	214523	216973: contig of 2453 bp in length
*	216974	217075: gap of unknown length
*	217076	219621: contig of 2546 bp in length
*	219622	219721: gap of unknown length
*	219722	221777: contig of 2056 bp in length
*	221778	221877: gap of unknown length
*	221878	222899: contig of 1022 bp in length
*	222899	222999: gap of unknown length
*	222900	224359: gap of unknown length
*	223000	224459: gap of unknown length
*	224360	224459: gap of unknown length
*	224460	225532: contig of 1073 bp in length.

FEATURES  
 source  
 1..225532  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /chromosome="Rf1"  
 /clone="CH230-59C15"  
 BASE COUNT 61510 a 52080 c 49274 g 60665 t 2003 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 10.3 Length: 225532  
 Score: 144.50 Matches: 156  
 Percent Similarity: 36.41% Conservative: 77  
 Best Local Similarity: 24.38% Mismatches: 213  
 Query Match: 3.88% Indels: 198  
 DB: 2 Gaps: 33

US-09-809-638-2 (1-699) x AC096364 (1-225532)  
 Oy 100 ValSerSerLeuIleValGlnAlaValThrTPPTPSPGlySerHis----- 116  
 DB 214157 GTCACACATCTTACTT-----CCAAATGAGTCAATGTCGTGTCAGAGTCTCTGCT 214104



JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
3 (bases 1 to 31717)  
Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Submitted (03-AUG-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB0 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
CDS are numbered using the following system eg SCJ21.01c. SC (S.  
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary  
strand).

The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.gov/jp/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or att) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid J21 lies between J11 and J33 on the AseI-J genomic  
restriction fragment.  
Location/Qualifiers  
1..31717  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid J21"  
complement(1..342)  
/gene="SCJ21.01c"  
1..136  
/note="nominal overlap with SCJ11"  
complement(1..342)  
/gene="SCJ21.01c"  
/note="SCJ21.01c, probable integral membrane protein,  
partial CDS, len: >114 aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative integral membrane protein"  
/protein\_id="CAB52347.1"  
/db\_xref="GI:5708234"  
/translation="MRRHLISFVARSVALIGHARNPAMLLVYVYIPWTAIVYIT  
PDQAPPRLRATGCVLSPPGNHLTIQTGALNAVTLIAGFMFAATFTGGADRLLAMA  
GYRPHLVLAAL"  
complement(339..953)  
/gene="SCJ21.02c"  
complement(339..953)  
/gene="SCJ21.02c"  
/note="SCJ21.02c, probable ABC transport system  
ATP-binding protein, len: 204 aa; similar to many e.g.  
MUC\_ECOLI high-affinity zinc uptake system ATP-binding  
protein (251 aa), fasta scores; opt: 327 z-score: 404.3  
E(): 3.6e-15, 35.2% identity in 179 aa overlap. Contains

FEATURES  
SOURCE

## gene

## misc\_feature

## CDS

gene  
CDS

PS00017 ATP/GTP-binding site motif A (P-loop) and Pfam  
match to entry PF00005 ABC\_tran, ABC transporter"  
/codon\_start=1  
/transl\_table=1  
/product="ABC transport system ATP-binding protein"  
/protein\_id="CAB52348.1"  
/db\_xref="GI:5708235"  
/translation="MDVHDVHNSYRQAVLRGVRLRPTKTIAGVNGAGKRTLLK  
ILSELPRDGTVYHNSRGRCYQTYVLDSPFYRQRLDPRFRSFGITGERRAEVVE  
TLAFETLDQAGRLSGSRKMLTALHNDPVELIDPEITGDFMETYLRTWEIAT  
RLNDAGRSYLVSLADYIDLQDLWRLIEGGRILIDERPMEAA"  
complement(350..354)  
/gene="SCJ21.02c"  
complement(381..878)  
/gene="SCJ21.02c"  
/note="Pfam match to entry PF00005 ABC\_tran, ABC  
transporter, score 147.80, E-value 1.9e-40"  
complement(834..857)  
/gene="SCJ21.02c"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement(961..964)  
/gene="SCJ21.03"  
/note="Pfam match to entry PF00743 FMO-like,  
flavin-binding monooxygenase-like, score 47.90, E-value  
7.5e-25"  
1194..2570  
/gene="SCJ21.03"  
1194..2570  
/note="SCJ21.03"  
/translation="MRRVCYIAGSPGLASRYLASRGIDPFDCYAGSGIGGLMRYGN  
DNGMSGVYASLHANISKESMSFSLPMPDSYPVPHHTOVLAYLESYAEFTGLHGTG  
LRTEVTSVRPVEGSGMEVTRRSRGAEEDRYEVVANGHMDPRLPPAVGAGY  
FEGSAVAHAHYRSPPEYAGORLVYVNGSGCEIAEISAAARTTSARAVAHFPRK  
MLGRPADHNAVSAAVLRPRKPGKALLRLRGAPALYATGYSLSPFLPAPVPS  
DELVLQARGAVTAKPGIRSGRDSVFTDGRRTVDVAVYATGYSLSPFLPAPVPS  
APDRTELVTATPPRLPGLEFGLAPAGAAFLPQAEWLADLIEGVELLPPE  
MTRSIARARERHDKVAPSYRHGIEIDICAYRRALRRELRRARRRADAARRDAPR  
PAPVSDTV"  
complement(2660..4006)  
/gene="SCJ21.04c"  
complement(2660..4006)  
/gene="SCJ21.04c"  
/note="SCJ21.04c, unknown, len: 448 aa; some similarity to  
M. tuberculosis hypotheical protein TR\_P71694  
(EMBL:295844) Rv1425 (459 aa), fasta scores; opt: 150  
z-score: 166.6 E(): 0.063, 23.9% identity in 464 aa  
overlap"  
/codon\_start=1  
/transl\_table=1  
/product="hypotheical protein SCJ2.04c"  
/protein\_id="CAB52350.1"  
/db\_xref="GI:5708237"  
/translation="MSAPPTAPADPSTGSDRPKRSVPPPHPTAVDRAFLDERRP  
EVRMDAGVAATLSGPPSLADRAIVAFRIAGELTLTSYTGAGPAPWQSDSEPDVR  
HVEVVAVDGAPDMWTALHSLANAPGAYGWVLIGHGEVADALVATRFIARFTVRRHDPDA  
AAATPFAMIGDGPSPALPGSGVGVPFRCAVALATRFIARFTVRRHDPDA  
FAPAGERLWRGVRVDPVTLRLIGAGRGSGAHDAHLALAGLALSWAARSGVPLPRVTA

Page 24

AL451104 Homo sapi  
AL451104 Homo sapi

# ALIGNMENTS

RESULT 1  
AK025164  
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL05748.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished (2000)  
REFERENCE 2 (bases 1 to 2486)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human

**THIS PAGE BLANK (00710)**

QY	1	ATGACCTCGCGTGGAGAGAAATCCCTTGGAGCGCGTGGGGGATGCTTTCTTGCTCT	60
Db	183	ATGACCTCGCGTGGAGAGAAATCCCTTGGAGCGCGTGGGGATGCTTTCTTGCTCT	24.2
QY	61	CTTACCATGACCTGGGACCGATGATCTATTACTCTTCTTGGCAACACATGAACTACT	120
Db	243	CTTACCATGACCGGGACCGATGATCTATTACTCTTCTTGGCAACACATGAACTACT	30.2
QY	121	GGGCTTAAGTTTAAATAGCATTTCTTCTCCAAATATTCCTAACAAATTAATCTCTTC	180
Db	303	GGGCTTAAGTTTAAATAGCATTTCTTCTCCAAATATTCCTAACAAATTAATCTCTTC	36.2
QY	181	TGGAAATGGTTTAAACAGAGTGGATGCTAAACCTGCTGAGGATATACATATTGGAGC	240
Db	363	TGGAAATGGTTTAAACAGAGTGGATGCTAAACCTGCTGAGGATATACATATTGGAGC	42.2
QY	241	ATAGCCCTCTCCACGCGCTCCAAATAGCCAAACTTCGACTGATGGTCTTGGCGTTGGCGTG	300
Db	423	ATAGCCCTCTCCACGCGCTCCAAATAGCCAAACTTCGACTGATGGTCTTGGCGTTGGCGTG	48.2
QY	301	TCTTCTCACTGATAGTGCAGAGCTGTGACTTGGTGCTGCGGAAGTCATTTGCCAAGGTAC	360
Db	483	TCTTCTCACTGATAGTGCAGAGCTGTGACTTGGTGCTGCGGAAGTCATTTGCCAAGGTAC	54.2
QY	361	CTCGAATTTGGGGATTCATTTTAGCAGACATTTGTTCTTCTTACGCATATGGTAT	420
Db	543	CTCGAATTTGGGGATTCATTTTAGCAGACATTTGTTCTTCTTACGCATATGGTAT	60.2
QY	421	ACTTCATTAACCAATCTGGAGTTACAGATGTCCAAAGATGATCTGACATTTAAGT	480

Db	603	ACTTCATTAACCCAAATCTGAGTTATTCAGATGTCCAAAGATATCTGACATTAACT	662
Qy	481	GCCTATACCACACTTGTATGCTATTGGCACAGATGCTGACTGCATGAATTAACCTGAAAGAAAG	540
Db	663	GCCTATACCACACTTGTATGCTATTGGCACAGATGCTGACTGCATGAATTAACCTGAAAGAAAG	722
Qy	541	AAGACTGTAGGTAGACGACGAGGGATGGCCTGTAGACCCAACTGGCTGTGGCAGGGCT	600
Db	723	AAGACTGTAGGTAGACGACGAGGGATGGCCTGTAGACCCAACTGGCTGTGGCAGGGCT	782
Qy	601	GCYTTTGTAGCCTTGTTCTCTACCCACATGGGTTTTTGAGAAAGTCTCTTGTATTCC	660
Db	783	GCYTTTGTAGCCTTGTTCTCTACCCACATGGGTTTTTGAGAAAGTCTCTTGTATTCC	842
Qy	661	AGATGGGACATGAGTGGGCAATCCAAATCCAGGGCCAGATATCCATTAACCATTTGGAGTGA	720
Db	843	AGATGGGACATGAGTGGGCAATCCAAATCCAGGGCCAGATATCCATTAACCATTTGGAGTGA	902
Qy	721	GTACTGCTGTGCTTGGCAATGATGTATGATGCTTCCATCTGTGTTGTGGTTGCTGTACT	780
Db	903	GTACTGCTGTGCTTGGCAATGATGTATGATGCTTCCATCTGTGTTGTGGTTGCTGTACT	962
Qy	781	GGTTTGATCTGGTGGGTACAGSAAACGCTTACGCTGCGGGCTCCTTTACTGCACACA	840
Db	963	GGTTTGATCTGGTGGGTACAGSAAACGCTTACGCTGCGGGCTCCTTTACTGCACACA	1022
Qy	841	TGGGCACTCTGTGCTGTGCTGTGCTTCCGCAATTTTACATGCATATGAGGGCCCAA	900
Db	1023	TGGGCACTCTGTGCTGTGCTGTGCTTCCGCAATTTTACATGCATATGAGGGCCCAA	1082
Qy	901	ACACTTGGACACTTATTAATCTAGGAGACAAACCTGGGAAAACATGACCATTTGCCATG	960
Db	1083	ACACTTGGACACTTATTAATCTAGGAGACAAACCTGGGAAAACATGACCATTTGCCATG	1142
Qy	961	ATATTTATCTCTAGAAAATTTTCTGAGCCTGGTGGACAGCTTAAAGTTGTCCCA	1020
Db	1143	ATATTTATCTCTAGAAAATTTTCTGAGCCTGGTGGACAGCTTAAAGTTGTCCCA	1202
Qy	1021	GGAGGTGTACGCTAGAGAAAGATCAGATGTGCTTTTGGGAGACATGATGTTAATTATC	1080
Db	1203	GGAGGTGTACGCTAGAGAAAGATCAGATGTGCTTTTGGGAGACATGATGTTAATTATC	1262
Qy	1081	GGGCTGAATATGCTATTTGGTCTCTAAGAAAACCTTGACTGCTCTTCAACAAAAC	1140
Db	1263	GGGCTGAATATGCTATTTGGTCTCTAAGAAAACCTTGACTGCTCTTCAACAAAAC	1322
Qy	1141	AGTCTCTAAATGCTTTTCAAAAAGTAGTGAATAATACATAAACTTTTCTGTGGCTGTT	1200
Db	1323	AGTCTCTAAATGCTTTTCAAAAAGTAGTGAATAATACATAAACTTTTCTGTGGCTGTT	1382
Qy	1201	GTTGGTGGGATTTGTTGGATTAAGACATACGGCATAAAGCCTATGAGAAAACCTGGC	1260
Db	1383	GTTGGTGGGATTTGTTGGATTAAGACATACGGCATAAAGCCTATGAGAAAACCTGGC	1442
Qy	1261	AAATGGCACCAACCAAAAGGCTCTGGGCGCATGTGGCCTTCAAGTTTGGATATATAC	1320
Db	1443	AAATGGCACCAACCAAAAGGCTCTGGGCGCATGTGGCCTTCAAGTTTGGATATATAC	1502
Qy	1321	AATGAAGGCTGTACTGTAGAAAGATCAGCTCACCTGCTCAATGAACACAGTGCAGAT	1380
Db	1503	AATGAAGGCTGTACTGTAGAAAGATCAGCTCACCTGCTCAATGAACACAGTGCAGAT	1562
Qy	1381	TTCTATACATTTTGGAGATGATGCTTCTAAGCCTATATGGGAAACAATGACTTAAAC	1440
Db	1563	TTCTATACATTTTGGAGATGATGCTTCTAAGCCTATATGGGAAACAATGACTTAAAC	1622
Qy	1441	ATGAGCTAAGGGGAAAGTTGCGTTTCTATACAGATTTGGTGCACAAGCAAGATATAC	1500
Db	1623	ATGAGCTAAGGGGAAAGTTGCGTTTCTATACAGATTTGGTGCACAAGCAAGATATAC	1682
Qy	1501	ACTTGGGGGATATGCTTTGTCAAGATACCAATTTGTGAATCTGAGATCACCCTTCT	1560
Db	1683	ACTTGGGGGATATATGCTTTGTCAAGATACCAATTTGTGAATCTGAGATCACCCTTCT	1742

**THIS PAGE BLANK (09710)**



ed Oct 23 09:09:13 2002

us-09-

Qy 1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTTCGGGCAAG 1620  
|||||  
Db 1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTTCGGGCAAG 1802  
Qy 1621 CTGGTGGATTTTGTCTGACACACTTTGGGAACACGAAGATGACCTCGACAGGAACTG 1680  
|||||  
Db 1803 CTGGTGGATTTTGTCTGACACACTTTGGGAACACGAAGATGACCTCGACAGGAACTG 1862  
Qy 1681 CAGGCTATTGCTGTTTCAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740  
|||||  
Db 1863 CAGGCTATTGCTGTTTCAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922  
Qy 1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800  
|||||  
Db 1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982  
Qy 1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860  
|||||  
Db 1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042  
Qy 1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAAGTGAATTCAGAAATTCAGATG 1920  
|||||  
Db 2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAAGTGAATTCAGAAATTCAGATG 2102  
Qy 1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980  
|||||  
Db 2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162  
Qy 1981 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAAATCCAGATTTGGATCCTACAAAGAA 2040  
|||||  
Db 2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAAATCCAGATTTGGATCCTACAAAGAA 2222  
Qy 2041 GGACACAATTATGAAAAACAACCATCATTTTCATATGAATACTCCCAATACTTTTATGA 2100  
|||||  
Db 2223 GGACACAATTATGAAAAACAACCATCATTTTCATATGAATACTCCCAATACTTTTATGA 2282  
Qy 2101 AAC 2103  
|||  
Db 2283 AAC 2285

RESULT 2  
BC016523  
LOCUS  
D

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:04:48 : Search time 3823 Seconds  
(without alignments)  
11511.511 Million cell updates/sec

Title: US-09-809-638-1  
Perfect score: 2103  
Sequence: 1 atgacgcgcgtgtgagaga.....ccaaactctttatgaac 2103

Scoring table: IDENTITY\_NUC  
Gapop 10.0', Gapext 1.0

Archived: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Geneml:\*

- 1: gb\_ba:\*
- 2: gb\_hcg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hcg\_hum:\*
- 31: em\_hcg\_inv:\*
- 32: em\_hcg\_other:\*
- 33: em\_hggo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query  
No. Score Match Length DB ID Description

Result No.	Query Score	Match	Length	DB ID	Description
1	2098.2	99.8	2486	9	AK025164
2	1511	71.8	2412	10	BC016523
3	267.4	12.7	184206	9	AC020593
4	160	7.6	222885	2	AL158816
5	151	7.2	222885	2	AL158816
6	132.4	6.3	31613	8	SCCHR111
7	121.8	5.8	1234	8	DB9130
8	121.8	5.8	34034	8	SPAC588
9	112	5.3	305	6	AX337478
10	110.6	5.3	995	11	CNS06JW
11	81.6	3.9	165893	2	AC106576
12	81.4	3.9	1348	9	HUMTPPSA
13	54	2.6	54	6	AX106346
14	54	2.6	54	6	AX140637
15	54	2.6	54	6	AX200497
16	54	2.6	54	6	AX267153
17	42.6	2.0	96374	2	AC095821
18	42.4	2.0	151931	30	AC011956
19	42	2.0	162349	9	AC006226
20	42	2.0	190000	2	AC007072
21	42	2.0	199528	2	AC108864
22	41.4	2.0	167750	2	AC016439
23	41.4	2.0	172710	2	AC021928
24	41.2	2.0	141577	9	AL606748
25	41	1.9	6563	10	RATCYP45OA
26	41	1.9	209712	2	AL627072
27	40.8	1.9	408	11	GS1484
28	40.8	1.9	42114	3	US0193
29	40.8	1.9	124067	9	AP000435
30	40.8	1.9	147358	9	AP002345
31	40.6	1.9	94947	2	AC099240
32	40.4	1.9	100061	2	AC016551
33	40.4	1.9	170569	2	AC108109
34	40.4	1.9	191132	9	AC010282
35	40.2	1.9	34568	2	AC083896
36	40.2	1.9	96280	9	AC006968
37	38	1.9	209524	2	AC099602
38	39.8	1.9	1141	6	AX083744
39	39.8	1.9	54814	2	AC104718
40	39.8	1.9	67460	2	AC102147
41	39.6	1.9	6917	10	RATCYP45C
42	39.4	1.9	20250	10	AF260315
43	39.4	1.9	163782	9	AC011228
44	39.4	1.9	177402	9	AC019210
45	39.2	1.9	71229	2	AL451104

## ALIGNMENTS

RESULT 1  
AK025164  
LOCUS  
DEFINITION Homo sapiens cDNA: FLJ21511 fls, clone COL05748.  
ACCESSION AK025164  
VERSION AK025164.1 GI:10437625  
KEYWORDS oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE  
1 (sites)  
Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project  
Unpublished (2000)  
2 (bases 1 to 2486)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.  
Direct Submission  
Submitted (29-AUG-2000) Sunto Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human

## COMMENT

Genome Center: Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing; Research Association for Biotechnology; cDNA Library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

## FEATURES

## Source

Location/Qualifiers  
 1..2486  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COL05748"  
 /tissue\_type="colon"  
 /clone\_lib="COL"  
 /note="cloning vector pME18SFL3"  
 183..2282  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB15080.1"  
 /db\_xref="GI:10437626"

## CDS

translation="WTSMLREILLESLLGCVSWSLYHDLGPMIYFPIQLTELEB  
 FSIAPLSPIFLTPFVKLVNKKMLTLRLITLIGSISFAPNKLRLMLALGVSS  
 SLIYQAVTMSGSHLQRLRWGLIQLIIVLWTLNWTSLPIMSYOMNVLIVLS  
 ALATDRIGTDDCSKPEPKTGEVATGASPNMLLGAAGSLVFLYHWGVSIL  
 VSRMAVGHPPGPDNPFGGAVLILCLASGLMLPCLMFRGGLIMFLYHWASAGIL  
 VLMMAAVSGCFEFTASMPOTLGLHINSGRPGKMTIAMFLYLEIFPCMCCT  
 AFKVPGVYIARESDVLLGLTMLILGLMLGPKKNDLLQTKNSKVLPRKSEKY  
 MKLFLMLVGLGLGLRHKAVERRIKVAPEKVSAAIMFERGYNEGWSLERS  
 AHLNLETAGFTLLESDSKPYKGNNDLTMWLGKLFYDGFSTYHTWGLIALS  
 RPYIKSEHLLSPGEIAPATLTIVNSGLVDFYHGEHEDDRLKGLAIVS  
 KLIKSSNOVIFLGITSAFSDVLIOLIEHGNVDIDSTHDMRCEYIMRGLIRLG  
 YARISHAELSDSEOMAKFRIPDDPTNRRDNCKYVIDHREVSERKIHENPRGSYKSH  
 YENNNHMHMTPKFEL"

BASE COUNT 692 a 516 c 606 g 672 t  
 ORIGIN

## Query Match

Best Local Similarity 99.98; Score 2098.2; DB 9; Length 2486;

Matches 2100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGACCTCGCTGAGAGAAATCTCTTGAGTCGCTCGTGGAGTGTCTTCTGGCT 60  
 183 ATGACCTCGCTGAGAGAAATCTCTTGAGTCGCTCGTGGAGTGTCTTCTGGCT 242  
 61 CTCTACATGACCTGGACCGATGATCTATTCTTCTTGGCAACCTAGACCTACT 120  
 243 CTCTACATGACCTGGACCGATGATCTATTCTTCTTGGCAACCTAGACCTACT 302  
 121 GGGCTGAAGGTTTGTATGATCTTCTTCCCAATATTCCTCAATTAATCTCTTC 180  
 303 GGGCTGAAGGTTTGTATGATCTTCTTCCCAATATTCCTCAATTAATCTCTTC 362  
 181 TGGAAATGTTAAACAAGATGATGTAACCTGCTGAGATAATCACTATTGGCAGC 240  
 363 TGGAAATGTTAAACAAGATGATGTAACCTGCTGAGATAATCACTATTGGCAGC 422  
 241 ATAGCCTCTTTCAGGCTCCAAATGCAACTTGCATGATGTTTGGCTTGGGGTG 300  
 423 ATAGCCTCTTTCAGGCTCCAAATGCAACTTGCATGATGTTTGGCTTGGGGTG 482  
 301 TCTTCCTCACTATAGTGAAGTGTGATGCTGTGGTGGGAAGTCAATTGCAAGGTAC 360  
 483 TCTTCCTCACTATAGTGAAGTGTGATGCTGTGGTGGGAAGTCAATTGCAAGGTAC 542  
 361 CTCAGAAATTTGGGATTTCAATTTAGACAGATGTTCTTGTCTTACGATATGGTAT 420  
 543 CTCAGAAATTTGGGATTTCAATTTAGACAGATGTTCTTGTCTTACGATATGGTAT 602  
 421 ACTTCACTAAACCAATCTGAGATTATCGATGTCCAAAGATGATCTGACATTAACT 480

603 ACTTCATAACCAATCTGAGATTATCGATGTCCAAAGATGATCTGACATTAACT 662  
 481 GCCATAGCCACACTGATGATGTAATTTGACAGATGAGTGCAGTAATCAAGAAAG 540  
 663 GCCATAGCCACACTGATGATGTAATTTGACAGATGAGTGCAGTAATCAAGAAAG 722  
 541 AAGACTGTGAGGTAGCCACAGGGGATGCGCTTAGACCCCACTGCTGCGAGGGCT 600  
 723 AAGACTGTGAGGTAGCCACAGGGGATGCGCTTAGACCCCACTGCTGCGAGGGCT 782  
 601 GCTTTTGGTAGCTGTCTCTTCTCAACCCCTGCGGTTTGGAGAAGTCTCTTGTTC 660  
 783 GCTTTTGGTAGCTGTCTCTTCTCAACCCCTGCGGTTTGGAGAAGTCTCTTGTTC 842  
 661 ACATGGGCGAGTAGTGGGATCCATCCAGGGGCCAGATCTTAACCATTTGGAGTGA 720  
 843 ACATGGGCGAGTAGTGGGATCCATCCAGGGGCCAGATCTTAACCATTTGGAGTGA 902  
 721 GTACTGCTGTGCTTGGCAAGTGGATGATCTTCATCTTGTGTGTTGCTGTACT 780  
 903 GTACTGCTGTGCTTGGCAAGTGGATGATCTTCATCTTGTGTGTTGCTGTACT 962  
 781 GGTTCATCTGCTGGGTACAGGAACAGCTTACGTCGGGGCTCTTACTGACACACA 840  
 963 GGTTCATCTGCTGGGTACAGGAACAGCTTACGTCGGGGCTCTTACTGACACACA 1022  
 841 TGGGCGAGCTGTGTGCTGCTGTGCTTGGCATCTTACTGATCCATGTCGCGCCAA 900  
 1023 TGGGCGAGCTGTGTGCTGCTGTGCTTGGCATCTTACTGATCCATGTCGCGCCAA 1082  
 901 ACACCTTGACACCTTATTAACACAGGCAAAACCTGGGAAAACCATGACCATTCAT 960  
 1083 ACACCTTGACACCTTATTAACACAGGCAAAACCTGGGAAAACCATGACCATTCAT 1142  
 961 AATTTTATCTCTAGAAATATTTTCTGCTGCTGCTGACAGCTTTTAAGTTGCCA 1020  
 1143 AATTTTATCTCTAGAAATATTTTCTGCTGCTGCTGACAGCTTTTAAGTTGCCA 1202  
 1021 GGAGGTGTCTACGATAGAGAAAGATCAGATGCTTTTGGGGCAATGATTAATATTC 1080  
 1203 GGAGGTGTCTACGATAGAGAAAGATCAGATGCTTTTGGGGCAATGATTAATATTC 1262  
 1081 GGGCTGAATATCTATTGCTCTTAAGAAAACCTTGACTGCTTCTTCAACCAAAAAC 1140  
 1263 GGGCTGAATATCTATTGCTCTTAAGAAAACCTTGACTGCTTCTTCAACCAAAAAC 1322  
 1141 AGTTCTAAGTCTTTTACAGAAAGATGAAAATACATGAACCTTTTCTGCTGCTCT 1200  
 1323 AGTTCTAAGTCTTTTACAGAAAGATGAAAATACATGAACCTTTTCTGCTGCTCT 1382  
 1201 GTTGTGTGGGATTTTGGGATTTAGACTACGCAATTAAGCCTATGAGAGAAACCTGGGC 1260  
 1383 GTTGTGTGGGATTTTGGGATTTAGACTACGCAATTAAGCCTATGAGAGAAACCTGGGC 1442  
 1261 AAAGTGGCACCAACAAAGAGTCTCTGCTGCTGCTTTCAGGTTTGGATATGAG 1320  
 1443 AAAGTGGCACCAACAAAGAGTCTCTGCTGCTGCTTTCAGGTTTGGATATGAG 1502  
 1321 AATGAAGGTGTGTAGCTTACAAAGATCAGTCACTGCTCAATGAAGAAAGTGTCAAT 1380  
 1503 AATGAAGGTGTGTAGCTTACAAAGATCAGTCACTGCTCAATGAAGAAAGTGTCAAT 1562  
 1381 TTTCATTAACAATTTTGGAGATGATCTTCAAGCCCTATATGGGGAACAATGACTTAAC 1440  
 1563 TTTCATTAACAATTTTGGAGATGATCTTCAAGCCCTATATGGGGAACAATGACTTAAC 1622  
 1441 ATGTGCTAAGGGGAAAGTGTGCTTCTATACAGACTTTGTCAGACCAAGATATGAG 1500  
 1623 ATGTGCTAAGGGGAAAGTGTGCTTCTATACAGACTTTGTCAGACCAAGATATGAG 1682  
 1501 ACTTGGGGGATATGCTTGTGCAAGATACCAATTTGGAATTCGACATCACCTCTT 1560  
 1683 ACTTGGGGGATATGCTTGTGCAAGATACCAATTTGGAATTCGACATCACCTCTT 1742

Oy 1561 CCGTCACGAGGGGAGATGCGACACCATGACATTGACGTTACATTTTCGGGCAAG 1620  
 Db 1743 CCGTCACGAGGGGAGATGCGACACCATGACATTGACGTTACATTTTCGGGCAAG 1802  
 Oy 1621 CTGTGTGATTTTGTGTCGACACACTTTTGGGAACACGAGATGACCTGACAGCAACTG 1680  
 Db 1803 CTGTGTGATTTTGTGTCGACACACTTTTGGGAACACGAGATGACCTGACAGCAACTG 1862  
 Oy 1681 CAGGCTATGCTGTTTCAAACTACTGAAAAGTACTCTTAATCAAGTATATTTCTGGGA 1740  
 Db 1863 CAGGCTATGCTGTTTCAAACTACTGAAAAGTACTCTTAATCAAGTATATTTCTGGGA 1922  
 Oy 1741 TATATCACTTACGACCGCGCCGACGATATATACAGCTCAGACATGCGCAATGTG 1800  
 Db 1923 TATATCACTTACGACCGCGCCGACGATATATACAGCTCAGACATGCGCAATGTG 1982  
 Db 1801 AAGATATCGACAGCAGCATGATCATGACAGATGCTGATACATATATGATGAGGCTG 1860  
 Db 1983 AAGATATCGACAGCAGCATGATCATGACAGATGCTGATACATATATGATGAGGCTG 2042  
 Oy 1861 ATCAGTTGGGTTATGCAAGAAATCTCCATGCTGACTGATGATTCAGAAATTCAGATG 1920  
 Db 2043 ATCAGTTGGGTTATGCAAGAAATCTCCATGCTGACTGATGATTCAGAAATTCAGATG 2102  
 Oy 1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTTATAGAGCAACCAAGTGGTCATA 1980  
 Db 2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTTATAGAGCAACCAAGTGGTCATA 2162  
 Oy 1981 GACCAAGAGAAATTTCTGAGAAATTTCAATTTATCCAGATTTGATCTTACAAAGAA 2040  
 Db 2163 GACCAAGAGAAATTTCTGAGAAATTTCAATTTATCCAGATTTGATCTTACAAAGAA 2222  
 Oy 2041 GGACCAATTTATGAACAACCATCATTTTCAATGAAATCTCCCAATCTTTTATGCA 2100  
 Db 2223 GGACCAATTTATGAACAACCATCATTTTCAATGAAATCTCCCAATCTTTTATGCA 2282  
 Oy 2101 AAC 2103  
 Db 2283 AAC 2285

RESULT 2  
 LOCUS BC016523 2412 bp mRNA linear ROD 05-NOV-2001  
 DEFINITION Mus musculus, similar to hypothetical protein FLJ21511, clone  
 MGC:27925 IMAGE:3584006, mRNA, complete cds.  
 ACCESSION BC016523  
 VERSION BC016523.1 GI:16741399  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2412)  
 STRAUSBERG, R.  
 DIRECT SUBMISSION  
 Submitted (31-OCT-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
 Villalon, D.K., Luna, R.A., Hale, S.M., Huljk, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.M., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 35 Row: K Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.  
 Location/Qualifiers  
 1..2412  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="MGC:27925 IMAGE:3584006"  
 /tissue\_type="Mammary tumor, Metallothionien-TGF alpha  
 model, 10 month old virgin mouse. Taken by biopsy."  
 /clone\_11b="NCI CGAP\_Mam1"  
 /lab\_host="DH10B"  
 /note="Vector: pcMV-Sport6"  
 113..2212

## CDS

/codon\_start=1  
 /product="Similar to hypothetical protein FLJ21511"  
 /protein\_id="AAH16523.1"  
 /db\_xref="GI:16741400"  
 /translation="MPGLMRAIALEPLIGVYSLSYHGLSPMIVYPPLOTDELGLEF  
 FCVAFSLPIILLTIPMLKIVKKWTLILRIVYGSIASPAPAKIRIWLALGVSS  
 SLIVQYTMWMSGLORYLKTWGLILGLVLLVIRYTSINLPISOMSRVLTLS  
 AVAVLDRTGIDGDRNPBGKPREVATGRYSLSWLLPGAFLSLLEPLTHIEVSL  
 VSRVAAGHPHDPDPNPFEGAVLLGSSGLMSLSSSSMLHDAGLAMMTGAASAMGL  
 YLRVMAAVSGCVLAFTGSMWPOVLGLVNSGNSSEAMATGAILVLOFPFCMCT  
 AFKFPVGVVARERSDYLTGIVITIGISLFGKKNIDFLQKNSPTLRCSERY  
 MKLILMLFVGGILGILRHRYEROLGRGAPVYSAALIMPEPGRDNEGMBLEERS  
 AQLKREKADPTITLESBASKPYIGNNDLTHMLEKGLFTDRPSTDRTHGIMVLS  
 RYPLVRSHHLPLSPBEGELAIPTMYNVSRLVDVYTHGNEHEDLDRLOLAIVS  
 KILNCSNOVYLFGLYITSEPSRDYIDLTRKGNVKDIDSSDGMWCEYIMYRGILRG  
 YARISHAELSDSEIOMAKFRIPDOPANRDNOKVIVIHQVPKNIHNPREGSYKEGH  
 NYENTHHPHMTPKYEV"  
 BASE COUNT 628 a 561 c 616 g 607 t  
 ORIGIN

Query Match 71.8%; Score 1511; DB 10; Length 2412;  
 Best Local Similarity 82.4%; Pred. No. 0;  
 Matches 1733; Conservative 0; Mismatches 370; Indels 0; Gaps 0;

Oy 1 ATGACCTGCTGTGAGAGAAATCTCTTGGAGTGCCTGCTGGGATGCTTCTTGCT 60  
 Db 113 ATGCCAGGCTGTGAGAGCAATTTGCGACCTTGGCTGCTGCTCTCTGCT 172  
 Oy 61 CTCTACATGACCTGGGAGCCGATGATCTATTCTTCTTCCAAACACTAGACCT 120  
 Db 173 CTCTACATGACCTGGGAGCCGATGATCTATTCTTCTTCCAAACACTAGACCT 232  
 Oy 121 GGGCTTGAAGTTTATGATACATTTCTTCTCCAAATATTCCTTAACATTTCT 180  
 Db 233 GGGCTGGAATTTTGTGTACCTTCTCTCGATACCTTCAACATTTCTCCACTG 292  
 Oy 181 TCGAAATTTGTTAAACAGAGTGTCTAACCTGCTGAGATATACATTTTGGCAGC 240  
 Db 293 TCGAAATTTGTTAAACAGAGTGTCTAACCTGCTGAGATATACATTTTGGCAGC 352  
 Oy 241 ATAGCTCTTCCAGAGCTCCAAATGCTGACAGATGATGCTTGGCTGGGGTG 300  
 Db 353 ATAGCTCTTCCAGAGCTCCAAATGCTGACAGATGATGCTTGGCTGGGGTG 412  
 Oy 301 TCTTCTACTGATGATGACCTGACCTTGTGCTGGGAGAGTCATTTGCAAGGTAC 360  
 Db 413 TCTTCTCTTGTATGTAAGTGTGACCTGCTGCTGAGAGAGATTCGAAGGTAC 472  
 Oy 361 CTCAGATTTTGGGATTTATTTTGAAGATTTTCTTGTCTTCTACGATATGATG 420  
 Db 473 CTCAGATTTTGGGATTTATTTTGAAGATTTTCTTGTCTTCTTCTGGAATATGATG 532  
 Oy 421 ACTTCACTAAACCAATCTGAGATTATCAGATGTCACAAAGTATGATGATTAAGT 480

Db 533 ACATATTAAACCAATTTGGAGCTATCAGATGTCACAGAGATGTCACATTTGANT 592  
 Oy 481 GCCATTAGCCACTTGATGCTATGAGCAGATGGAGTACTGATTAACCTTAAGAAAG 540  
 Db 593 GCCGTAAGCTGCTGTGACGGAATGGACAGATGGTACTACAGAAACCTTAAGGGAAG 652  
 Oy 541 AAGACTGTGAGGTAGCCACAGGGGATGGCTCTAGACCACACTGGCTGCTGGAGGGCT 600  
 Db 653 AAGCCCGGTAGAGGTGGCCACGGGAGGAGCTCTATCTATGATGTTGGCTGGACGGCGC 712  
 Oy 601 GCTTTTGGAGCTGTGTCTCTACACCCACTGGCTTTTGGAGAACTCTCTTTGTTCC 660  
 Db 713 GCTTTGGAGGCTGCTGTTCTTAACCTGATCTTGGAGAGTCTCCATTGTTTCC 772  
 Oy 661 ACATGGGAGTGTGAGGATCCACATCCAGGGCCAGATCCTAACCCATTTGAGGTGCA 720  
 Db 773 AGATGGGAGTGTGAGGATCCACATCCAGGGCCAGATCCTAACCTTTTGGAGGTGCG 832  
 Oy 721 GTACTGCTGTGCTGGCAGTGGATGTGATCTTCATCTGTTTGGTGTGCTGCTACT 780  
 Db 833 GTTCTACTGTGCTGTCTCAAGTGAGTGTGATGCTATCGGGTTTCATCATGGCTCATGCT 892  
 Oy 781 GCTTTGATCTGTGGGTACAGAACAGTTCAGCTGGGGCTCTTACTGACACACA 840  
 Db 893 GGGTTAGCTGTGGATGACAGAGACGCTTGGCCATGGGCTCTCTACCTGCGCACA 952  
 Oy 841 TGGGAGCTGTGTGCTGTGCTGTCTGCTTCCGATCTTACTGATCCATTCATGGGCCCA 900  
 Db 953 TGGGAGCTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1012  
 Oy 901 ACATTTGACACCTTATTAATCTCAGGGGACAAACCTGGGAAACCATGACCTATTCATG 960  
 Db 1013 GTACTTGGCATCTTGTGAACTCAGGAAACAACTCGGAGAACCATGGCCACTGGCATG 1072  
 Oy 961 ATATTTTATCTTCTAGAAATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 Db 1073 ATCTTTATGTTTACAAACATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1132  
 Oy 1021 GGAGGTGCTACGTAGAGAAAGATCAGATGCTTTTGGGAGCAATGATGTTAATTATC 1080  
 Db 1133 GGAGGTGCTACGTAGAGAAAGATCAGATGCTTTTGGGAGCAATGATGTTAATTATC 1192  
 Oy 1081 GGGCTGAATATGCTATTTGGTCTTAAGAAAACTTGACTTCTCTCAACAAAAAAC 1140  
 Db 1193 GGATTTGATATGCTGTTGGACCTTAAGAAAACTTGACTTCTCTCAACAAAAAAC 1252  
 Oy 1141 AGTCTTAAGTCTTTTCAAGAAAGATGAAATATGAAATCTTTCTGCTGCTGCT 1200  
 Db 1253 ACTCTTAAGACCTTTTGAATGATGAGTAAATATGAAATCTATCTGCTGCTGCT 1312  
 Oy 1201 GTTGTGTGGAATTTGGGATTTAGACTACGCAATAAAGCCTATGAGAGAAATGAGGC 1260  
 Db 1313 GTTGTGTGGAATTTGGGATTTAGACTACGCAATAAAGCCTATGAGAGAAATGAGGC 1372  
 Oy 1261 AAAGTGGACCAACAAAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Db 1373 CGAGGGGACACCGCACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1432  
 Oy 1321 AATGAAGGTGTGATGCTTAAGATCAGTCACTGCTCAATGAAACAGTGCATATG 1380  
 Db 1433 AATGAAGGTGTGCTTAAGATCAGTCACTGCTCAATGAAACAGTGCATATG 1492  
 Oy 1381 TTCAATAAATTTTGGAGTGTCTTAAGCCTATGATGGGAAACATGATTAAC 1440  
 Db 1493 TTTATACACATTTTGGAGTGTCTTAAGCCTATGATGGGAAACATGATTAAC 1552  
 Oy 1441 ATGTGGTGAAGGAAAGTGGTCTTATACAGACTTGTGTCAGAGCAAGATATAC 1500  
 Db 1553 ATGTGGTGAAGGAAAGTGGTCTTATACAGACTTGTGTCAGAGCAAGATATAC 1612  
 Oy 1501 ACTTGGGGAATTTATGCTTTGTCAAGATACCAATTTGAAATCTGAGCATCCTTCT 1560

Db 1613 ACCTGGGGGATTTATGCTGTCTGCTGATACCCGATTTGTCAGATTCGAACATCCTTCT 1672  
 Oy 1561 CCGTACCAGAGGGGAGATGTCACACAGCATCATATTGACCTTAACTTTGGGCAAG 1620  
 Db 1673 CCGTCCAGAGGGGAGATGTCACACAGCATCATATTGACCTTAACTTTGGGCAAG 1732  
 Oy 1621 CCGTGGATTTTGTGCTGACACACTTTTGGAAACAGAGATGACCTGACAGAACTG 1680  
 Db 1733 CTGTGGAATTTTGTGCTGACACACTTTTGGAAACAGAGATGACCTGACAGAACTG 1792  
 Oy 1681 CAGGCTATTGCTGTTTCAAACTACTGAAAGTACTTATCAAGTATATTTTGGGA 1740  
 Db 1793 CAGGCTATTGCAAGTTTCAAACTGCTGAAAAATTTTCAAGTATATTTTGGGT 1852  
 Oy 1741 TATATCACTTACAGCTGCTGCTGAGATTTATACAGCTTCACTGAAATGCAATGTG 1800  
 Db 1853 TATATCACTTACAGCTGCTGCTGAGATTTATACAGCTTCACTGAAATGCAATGTG 1912  
 Oy 1801 AAGATATTCAGACATGATCAGATGAGATGCTGAAATATGATTTATGAGGGCTG 1860  
 Db 1913 AAGATATTTGACAGCTGATGAGATGAGATGCTGAGTATCATCATGATACCGGGCTG 1972  
 Oy 1861 ATCAGTGGGTATGACAAAGATCTCCATGCTGAACTGAGTATGATTCAGAAATTCAG 1920  
 Db 1973 ATCAGTGGGTATGACAAAGATCTCCATGCTGAACTGAGTATGATTCAGAAATTCAG 2032  
 Oy 1921 GCAAAATTTAGATCCCTGATGACCCCACTAATTTATGAGAACACAAAGTGTGCTA 1980  
 Db 2033 GCAAAATTTAGATCCCTGATGACCCCACTAATTTATGAGAACACAAAGTGTGCTA 2092  
 Oy 1981 GACACAGAGAAATTTGAGAAAAATTTAATTTAATTTAATTTGATTTGATTTGATTT 2040  
 Db 2093 GACACAGAGAAATTTGAGAAAAATTTAATTTAATTTAATTTGATTTGATTTGATTT 2152  
 Oy 2041 GGACACATTTATGAAACACATCATTTTATGATGATGATGATGATGATGATGATGAT 2100  
 Db 2153 GGACACATTTATGAAACACATCATTTTATGATGATGATGATGATGATGATGATGAT 2212  
 Oy 2101 AAC 2103  
 Db 2213 ACC 2215

RESULT 3  
 AC020593 184206 bp DNA linear PRI 07-NOV-2001  
 LOCUS Homo sapiens BAC clone RP11-317G22 from 4, complete sequence.  
 DEFINITION AC020593  
 ACCESSION AC020593.6 GI:14589709  
 VERSION hfg.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE  
 AUTHORS Swearingen, S. and Cotton, M.  
 TITLE The sequence of Homo sapiens BAC clone RP11-317G22  
 JOURNAL Unpublished  
 REFERENCE  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington  
 MO 63108, USA  
 REFERENCE  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 184206)  
 Waterston, R.  
 Direct Submission  
 Submitted (07-NOV-2001) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jul 3, 2001 this sequence version replaced g1:12057004.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0317622

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male  
 donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frengen, E.,  
 Tateo, M., Catanesse, J. J., and de Jong, P. U. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBAC3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is AC011956. Actual start of this  
 clone is at base position 1 of RPI1-317622; actual end is at base  
 position 184206 of RPI1-317622.

The sequence fidelity of RPI1-317622 between bases 17725 to 17956  
 can not be guaranteed due to an unresolved dinucleotide repeat. The  
 sequence fidelity of RPI1-317622 between bases 86965 to 87015 can  
 not be guaranteed due to an unresolved homopolymeric run.

## FEATURES

source  
 1. 184206  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RPI1-317622"  
 /clone\_11b="RPCI-11"  
 5. 70  
 repeat\_region  
 /rpt\_family="12"  
 252. 374  
 /note="match to EST BG169511 (NID:g12676214)"  
 misc\_feature  
 252. 374  
 /note="match to EST A1088656 (NID:g3427715) qb14a11.x1"  
 misc\_feature  
 252. 374  
 /note="match to EST BG214746 (NID:g13740767)"  
 252. 374  
 /note="match to EST BF055447 (NID:g10809343)"

misc\_feature  
 252. 374  
 /note="match to EST A1077951 (NID:g3412359) oy39a10.s1"  
 252. 374  
 /note="match to EST AM006753 (NID:g5855531) wt07a02.x1"  
 252. 374  
 /note="match to EST BG291592 (NID:g13049843)"  
 misc\_feature  
 252. 374  
 /note="match to EST W80356 (NID:g1391393) zh49a05.s1"  
 252. 494  
 /note="match to EST AA464688 (NID:g2189572) zx82a02.s1"  
 /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"  
 252. 374  
 /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"  
 252. 374  
 /note="match to EST BG165990 (NID:g12672693)"  
 256. 374  
 /note="similar to Bos taurus EST AV609860 (NID:g9745530)"  
 280. 374  
 /note="match to EST AW371794 (NID:g6876448)"  
 347. 374  
 /note="match to EST AA303948 (NID:g1956281)"  
 708. 1203  
 /note="match to EST BI063872 (NID:g14471399)"  
 1021. 1308  
 /note="match to EST T06930 (NID:g318079)"  
 1119. 1143  
 /note="match to EST W68728 (NID:g1377608) z037c01.s1"  
 1119. 1143  
 /note="match to EST A1220135 (NID:g3802338) q144g05.x1"  
 /note="match to EST A1220135 (NID:g3802338) q144g05.x1"  
 1119. 1143  
 /note="match to EST A1271747 (NID:g3890914) qj77b10.x1"  
 /note="match to EST BG167015 (NID:g12673728)"  
 1183. 1538  
 /note="match to EST BG546183 (NID:g13544848)"  
 1319. 1350  
 /rpt\_family="AT-rich"  
 1349. 1466  
 repeat\_region  
 /rpt\_family="Alu"  
 1490. 1538  
 /note="match to EST AA847380 (NID:g2933898) oF07d05.s1"  
 1490. 1538  
 /note="match to EST A1220135 (NID:g3802338) q144g05.x1"  
 1490. 1538  
 /note="match to EST W68728 (NID:g1377608) z037c01.s1"  
 1490. 1538  
 /note="match to EST BG167015 (NID:g12673728)"  
 1490. 1538  
 /note="match to EST A1271747 (NID:g3890914) qj77b10.x1"  
 1491. 1538  
 /note="match to EST W80356 (NID:g1391393) zh49a05.s1"  
 1491. 1538  
 /note="match to EST A1088656 (NID:g3427715) qb14a11.x1"  
 1491. 1508  
 /note="match to EST BG291592 (NID:g13049843)"  
 1491. 1538  
 /note="similar to Homo sapiens EST BG207366  
 (NID:g13729053)"  
 1491. 1538  
 /note="match to EST BG165990 (NID:g12672693)"  
 1491. 1538  
 /note="match to EST AA303948 (NID:g1956281)"  
 1491. 1538  
 /note="match to EST AW371794 (NID:g6876448)"  
 1491. 1538  
 /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"  
 1491. 1538  
 /note="similar to Bos taurus EST AV609860 (NID:g9745530)"  
 1491. 1538  
 /note="match to EST A1077951 (NID:g3412359) oy39a10.s1"  
 1491. 1538  
 /note="match to EST BG214746 (NID:g13740767)"  
 1491. 1538  
 /note="match to EST AM006753 (NID:g5855531) wt07a02.x1"

```
misc_feature      1491..1538      /note="match to EST BC169511 (NID:912676214)"
misc_feature      1491..1546      /note="match to EST BF055447 (NID:910809343)"
repeat_region     1561..1606      /rpt_family="AT-rich"
repeat_region     1696..1986      /rpt_family="Alu"
repeat_region     2175..2217      /rpt_family="Alu"
repeat_region     2311..2780      /rpt_family="L1"
repeat_region     2828..3176      /rpt_family="MaLR"
repeat_region     3217..3544      /rpt_family="MaLR"
repeat_region     3617..3648      /rpt_family="Alu"
repeat_region     3793..4078      /rpt_family="AT-rich"
repeat_region     4138..4267      /rpt_family="Alu"
repeat_region     4462..4655      /rpt_family="Alu"
repeat_region     4639..4688      /rpt_family="MERL-type"
misc_feature      /note="match to EST AW005009 (NID:95853713) wv02e08.x1"

Query Match      12.7%: Score 267.4; DB 9; Length 184206;
Best Local Similarity 94.5%; Pred. No. 7.6e-64;
Matches 277; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 799 ACAGAAACAGCTTACCTGCGGGGCTCCTTACCTGACACATGAGGCGAGCTGCTGCT 858
      |||||
DB 111216 ACAGAAACAGCTTACCTGCGGGGCTCCTTACCTGACACATGAGGCGAGCTGCTGCT 111275

QY 859 GCGTGTCTTGCGCCATCTTTACTGATCATGTCGCCCCAAACACTTGACACCTTATT 918
      |||||
DB 111276 GCGTGTCTTGCGCCATCTTTACTGATCATGTCGCCCCAAACACTTGACACCTTATT 111335

QY 919 AACTCGAGGACAAACCTGGGAAACCATGACCATTCATGATTTATTTATCTCTAGAA 978
      |||||
DB 111336 AACTCGAGGACAAACCTGGGAAACCATGACCATTCATGATTTATTTATCTCTAGAA 111395

QY 979 AATATTTCTTGCTGCTGTGACAGCTTTTAAGTTTGTCCAGAGGTGTCTACGCTAGA 1038
      |||||
DB 111396 AATATTTCTTGCTGCTGTGACAGCTTTTAAGTTTGTCCAGAGGTGTCTACGCTAGA 111455

QY 1039 GAAAGATCAGATGTCTTTGGGACATGATGTTAATTATCGGGCTGAATAT 1091
      |||||
DB 111456 GAAAGATCAGATGTCTTTGGGACATGATGTTAATTATCGGGCTGAATAT 111508

RESULT 4
AL158816 222885 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
DEFINITION AL158816
ACCESSION AL158816
VERSION AL158816.11 GI:9943990
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
            1 (bases 1 to 222885)
REFERENCE Sims,S.
            Direct Submission
            Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
COMMENT On Aug 29, 2000 this sequence version replaced gi:9926474.
            ----- Genome Center
            Center: Sanger Centre
```

```
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: d4105b4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 204390 bases at least Q40
Consensus quality: 215491 bases at least Q30
Insert size: 219685; sum-of-contigs
Insert coverage: 126549; 14.0% error; agarose-fp
Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
coverage: 6.72x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3599: contig of 3599 bp in length
* 3600 3699: gap of 100 bp
* 3700 12818: contig of 9119 bp in length
* 12819 12918: gap of 100 bp
* 12919 16398: contig of 3480 bp in length
* 16399 16498: gap of 100 bp
* 16499 18590: contig of 2092 bp in length
* 18591 18690: gap of 100 bp
* 18691 22372: contig of 3682 bp in length
* 22373 22472: gap of 100 bp
* 22473 26905: contig of 4433 bp in length
* 26906 27005: gap of 100 bp
* 27006 29581: contig of 2576 bp in length
* 29582 29681: gap of 100 bp
* 29682 33446: contig of 3765 bp in length
* 33447 33546: gap of 100 bp
* 33547 48536: contig of 14990 bp in length
* 48537 48636: gap of 100 bp
* 48637 55599: contig of 6963 bp in length
* 55600 55699: gap of 100 bp
* 55700 65124: contig of 9425 bp in length
* 65125 65224: gap of 100 bp
* 65225 69090: contig of 3866 bp in length
* 69091 69190: gap of 100 bp
* 69191 73894: contig of 4704 bp in length
* 73895 73994: gap of 100 bp
* 73995 82665: contig of 8671 bp in length
* 82666 82765: gap of 100 bp
* 82766 87239: contig of 4474 bp in length
* 87240 87339: gap of 100 bp
* 87340 93671: contig of 6332 bp in length
* 93672 93771: gap of 100 bp
* 93772 97550: contig of 3779 bp in length
* 97551 97650: gap of 100 bp
* 97651 100101: contig of 2451 bp in length
* 100102 100201: gap of 100 bp
* 100202 103416: contig of 3215 bp in length
* 103417 103516: gap of 100 bp
* 103517 107564: contig of 4048 bp in length
* 107565 107664: gap of 100 bp
* 107665 131778: contig of 24114 bp in length
* 131779 131878: gap of 100 bp
* 131879 135261: contig of 3383 bp in length
* 135262 135361: gap of 100 bp
* 135362 137913: contig of 2552 bp in length
* 137914 138013: gap of 100 bp
* 138014 140949: contig of 2936 bp in length
* 140950 141049: gap of 100 bp
```



```

* 141050 156466: contig of 15417 bp in length
* 156467 156566: gap of 100 bp
* 156567 164836: contig of 8270 bp in length
* 164837 164936: gap of 100 bp
* 164937 168950: contig of 4014 bp in length
* 168951 169050: gap of 100 bp
* 169051 173724: contig of 4674 bp in length
* 173725 173824: gap of 100 bp
* 173825 179991: contig of 6167 bp in length
* 179992 180091: gap of 100 bp
* 180092 190208: contig of 10117 bp in length
* 190209 190308: gap of 100 bp
* 190309 193471: contig of 3163 bp in length
* 193472 193571: gap of 100 bp
* 193572 197829: contig of 4258 bp in length
* 197830 197929: gap of 100 bp
* 197930 222885: contig of 24956 bp in length.
Location/Qualifiers
1. .222885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="RPCI-3"
1. .3599
/note="assembly_fragment:00027"
misc_feature
3700. .12818
/note="assembly_fragment:00056"
misc_feature
12919. .16398
/note="assembly_fragment:00093"
misc_feature
16499. .18590
/note="assembly_fragment:00125"
misc_feature
18691. .22372
/note="assembly_fragment:00140"
misc_feature
22473. .26905
/note="assembly_fragment:00167"
misc_feature
27006. .29581
/note="assembly_fragment:00175"
misc_feature
29682. .33446
/note="assembly_fragment:00178"
misc_feature
33547. .48536
/note="assembly_fragment:00416"
misc_feature
48637. .55599
/note="assembly_fragment:00542"
misc_feature
55700. .65124
/note="assembly_fragment:00565"
misc_feature
65225. .69090
/note="assembly_fragment:00583"
misc_feature
69191. .73894
/note="assembly_fragment:00649"
misc_feature
73995. .82665
/note="assembly_fragment:00709"
misc_feature
82766. .87239
/note="assembly_fragment:00789"
misc_feature
87340. .93671
/note="assembly_fragment:00874"
misc_feature
93772. .97550
/note="assembly_fragment:00924"
misc_feature
97651. .100101
/note="assembly_fragment:01052"
misc_feature
100202. .103416
/note="assembly_fragment:01082"
misc_feature
103517. .107564
/note="assembly_fragment:01216"
misc_feature
107665. .131778
/note="assembly_fragment:01234"
misc_feature
131879. .135261
/note="assembly_fragment:01280"
misc_feature
135362. .137913
/note="assembly_fragment:01337"
misc_feature
138014. .140949
/note="assembly_fragment:01643"
misc_feature
141050. .156466

```

```

/note="assembly_fragment:01660"
156567. .164836
/note="assembly_fragment:01690"
164937. .168950
/note="assembly_fragment:01692.0"
169051. .173724
/note="assembly_fragment:01748"
173825. .179991
/note="assembly_fragment:02061"
180092. .190208
/note="assembly_fragment:02140"
190309. .193471
/note="assembly_fragment:02150"
193572. .197829
/note="assembly_fragment:02181"
197930. .222885
/note="assembly_fragment:00834
clone_end:17
vector_side:right"
BASE COUNT 69666 a 41364 c 44947 g 63682 t 3226 others
ORIGIN

```

```

Query Match 7.6%, Score 160; DB 2; Length 222885;
Best Local Similarity 89.6%; Pred. No. 1.4e-33;
Matches 172; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```

```

OY 1863 CAGGTGGGTTATGCAAGATCTCCATGCTGAGTGTGATTCAGAAATTCAGATGCG 1922
|||||
DB 157674 CAGGTGGGTTATGCAAGATCTCCATGCTGAGTGTGATTCAGAAATTCAGATGCG 157733
|||||
OY 1923 AAAATTTAGATCCCTGATGAGCCACGTAATATGAGCAACGAAAGTGCTATGA 1982
|||||
DB 157734 AAAATTTAGATCCCTGATGAGCCACGTAATATGAGCAACGAAAGTGCTATGA 157793
|||||
OY 1983 CCACAGAGAGTTTCTGAGAAATTCATTTAATCCAGATTGAGATCTCAAGAAAGG 2042
|||||
DB 157794 CCACAGAGAGTTTCTGAGAAATTCATTTAATCCAGATTGAGATCTCAAGAAAGG 157853
|||||
OY 2043 ACACAATTATGA 2054
|||||
DB 157854 TTCAGATTATGA 157865
|||||

```

```

RESUME 5
AL158816/c 222885 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
DEFINITION
ACCESSION AL158816
VERSION AL158816.11 GI:9943990
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 222885)
REFERENCE
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 29, 2000 this sequence version replaced gi:9926474.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: d1405B4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

Consensus quality: 204390 bases at least Q40  
Consensus quality: 215527 bases at least Q30  
Consensus quality: 215491 bases at least Q20  
Insert size: 219685; sum-of-contigs  
Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality  
coverage: 6.72x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
consists of 33 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 3599: contig of 3599 bp in length  
3600 3699: gap of 100 bp  
12819 12818: contig of 9119 bp in length  
12819 12918: gap of 100 bp  
12819 16398: contig of 3480 bp in length  
16399 16498: gap of 100 bp  
16499 18590: contig of 2092 bp in length  
18591 18690: gap of 100 bp  
18691 22372: contig of 3682 bp in length  
22373 22472: gap of 100 bp  
22473 26905: contig of 4433 bp in length  
26906 27005: gap of 100 bp  
27006 29581: contig of 2576 bp in length  
29582 33446: contig of 3765 bp in length  
33447 33546: gap of 100 bp  
33547 48536: contig of 14990 bp in length  
48537 48636: gap of 100 bp  
48637 55599: contig of 6963 bp in length  
55600 55699: gap of 100 bp  
55700 65124: contig of 9425 bp in length  
65125 65224: gap of 100 bp  
65225 69090: contig of 3866 bp in length  
69091 69190: gap of 100 bp  
69191 73894: contig of 4704 bp in length  
73895 73994: gap of 100 bp  
73995 82665: contig of 8671 bp in length  
82666 82765: gap of 100 bp  
82766 87239: contig of 4474 bp in length  
87240 87339: gap of 100 bp  
87340 93671: contig of 6332 bp in length  
93672 93771: gap of 100 bp  
93772 97550: contig of 3779 bp in length  
97551 97650: gap of 100 bp  
97651 100101: contig of 2451 bp in length  
100102 100201: gap of 100 bp  
100202 103416: contig of 3215 bp in length  
103417 103516: gap of 100 bp  
103517 107564: contig of 4048 bp in length  
107565 107664: gap of 100 bp  
107665 131778: contig of 24114 bp in length  
131779 131878: gap of 100 bp  
131879 135261: contig of 3383 bp in length  
135262 135361: gap of 100 bp  
135362 137913: contig of 2552 bp in length  
137914 138013: gap of 100 bp  
138014 140949: contig of 2936 bp in length  
140950 141049: gap of 100 bp  
141050 156466: contig of 15417 bp in length  
156467 156566: gap of 100 bp  
156567 164836: contig of 8270 bp in length  
164837 164936: gap of 100 bp  
164937 168950: contig of 4014 bp in length  
168951 169050: gap of 100 bp  
169051 173724: contig of 4674 bp in length  
173725 173824: gap of 100 bp  
173825 179991: contig of 6167 bp in length

FEATURES  
source

179992 180091: gap of 100 bp  
180092 190208: contig of 10117 bp in length  
190209 190308: gap of 100 bp  
190309 193471: contig of 3163 bp in length  
193472 193571: gap of 100 bp  
193572 197829: contig of 4228 bp in length  
197830 197929: gap of 100 bp  
197930 222885: contig of 24956 bp in length.  
Location/Qualifiers  
1..222885  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP3-405B4"  
/clone\_1fb="RPC1-3"  
1..3599  
/note="assembly-fragment:00027"  
3700..12818  
/note="assembly-fragment:00056"  
12919..16398  
/note="assembly-fragment:00093"  
16499..18590  
/note="assembly-fragment:00125"  
18691..22372  
/note="assembly-fragment:00140"  
22473..26905  
/note="assembly-fragment:00167"  
27006..29581  
/note="assembly-fragment:00175"  
29682..33446  
/note="assembly-fragment:00178"  
33547..48536  
/note="assembly-fragment:00416"  
48637..55599  
/note="assembly-fragment:00542"  
55700..65124  
/note="assembly-fragment:00565"  
65225..69090  
/note="assembly-fragment:00583"  
69191..73894  
/note="assembly-fragment:00649"  
73995..82665  
/note="assembly-fragment:00709"  
82766..87239  
/note="assembly-fragment:00789"  
87340..93671  
/note="assembly-fragment:00874"  
93772..97550  
/note="assembly-fragment:00924"  
97651..100101  
/note="assembly-fragment:01052"  
100202..103416  
/note="assembly-fragment:01082"  
103517..107564  
/note="assembly-fragment:01216"  
107665..131778  
/note="assembly-fragment:01234"  
131879..135261  
/note="assembly-fragment:01280"  
135362..137913  
/note="assembly-fragment:01337"  
138014..140949  
/note="assembly-fragment:01643"  
141050..156466  
/note="assembly-fragment:01660"  
156567..164836  
/note="assembly-fragment:01690"  
164937..168950  
/note="assembly-fragment:01692.0"  
169051..173724  
/note="assembly-fragment:01748"  
173825..179991  
/note="assembly-fragment:02061"

```

misc_feature      180092..190208
                  /note="assembly_fragment:02140"
misc_feature      190309..193471
                  /note="assembly_fragment:02150"
misc_feature      193572..197829
                  /note="assembly_fragment:02181"
misc_feature      197930..222885
                  /note="assembly_fragment:00834
                  clone_end:77
                  vector_side:right"
BASE COUNT      69666 a 41364 c 44947 g 63682 t 3226 others
ORIGIN
Query Match      7.2% Score 151; DB 2; Length 222885;
Best Local Similarity 100.0%; Pred. No. 5e-31;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1508 GGATTATGCGCTTGTCAAGATACCAATTTGAAATCTGAGCATCACCTTTCGCTAC 1567
|||||
DB 215729 GGATTATGCGCTTGTCAAGATACCAATTTGAAATCTGAGCATCACCTTTCGCTAC 215670
|||||
QY 1568 CAGAGGGGAGATCGCACCAGCATCATTCACGCTTACATTTTCGGCAAGCTGTGG 1627
|||||
DB 215669 CAGAGGGGAGATCGCACCAGCATCATTCACGCTTACATTTTCGGCAAGCTGTGG 215610
|||||
QY 1628 ATTTGTCGTGACACACTTTGGGAACACGA 1658
|||||
DB 215609 ATTTGTCGTGACACACTTTGGGAACACGA 215579
|||||

RESULT 6
LOCUS      SCCHRIII 316613 bp DNA linear PLN 29-JUN-2001
DEFINITION S.cerevisiae chromosome III complete DNA sequence.
ACCESSION  X59720.543845 S49180 S58084 S93798
VERSION     X59720.2 GI:14588895
KEYWORDS    chromosome.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE   1 (bases 1 to 316613)
            Rad,M.R., Lutzentrup,K., Xu,G., Kleinhaus,U. and Hollenberg,C.P.
            The complete sequence of a 11,953 bp fragment from CIG on
            chromosome III encompasses four new open reading frames
            yeast 7 (5), 533-538 (1991)
JOURNAL     91377317
MEDLINE     2 (bases 1 to 316613)
REFERENCE   Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and
            Frontali,L.
            A putative serine/threonine protein kinase gene on chromosome III
            of Saccharomyces cerevisiae
JOURNAL     92254506
MEDLINE     3 (bases 1 to 316613)
REFERENCE   Biteau,N., Fremaux,C., Hebrard,S., Menara,A., Aigle,M. and
            Crouzet,M.
            The complete sequence of a 10.8kb fragment to the right of the
            chromosome III centromere of Saccharomyces cerevisiae
JOURNAL     92254505
MEDLINE     4 (bases 1 to 316613)
REFERENCE   Benit,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
            Sequence of the Sup1-RAD18 region on chromosome III of
            Saccharomyces cerevisiae
JOURNAL     92221691
MEDLINE     5 (bases 1 to 316613)
REFERENCE   Bolle,P.A., Giliquet,V., Berben,G., Dumont,J. and Hilger,F.
            The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
            CRY1 on chromosome III, reveals the presence of seven open reading
            frames
JOURNAL     Yeast 8 (3), 205-213 (1992)

```

```

MEDLINE 92245758
REFERENCE 6 (bases 1 to 316613)
AUTHORS  Sor,F., Charet,G., Fabre,F., Faye,G. and Fukuhara,H.
TITLE    Sequence of the HMR region on chromosome III of Saccharomyces
JOURNAL  cerevisiae
MEDLINE  Yeast 8 (3), 215-222 (1992)
REFERENCE 92245759
AUTHORS  7 (bases 1 to 316613)
TITLE    Skala,J., Purnelle,B. and Goffeau,A.
JOURNAL  The complete sequence of a 10.8 Kb segment distal of SUP2 on the
MEDLINE  right arm of chromosome III from Saccharomyces cerevisiae reveals
REFERENCE  seven open reading frames including the RVS16L, ADP1 and PGK genes
AUTHORS  Yeast 8 (5), 409-417 (1992)
TITLE    8 (bases 1 to 316613)
JOURNAL  Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M.,
MEDLINE  Alberghina,L., Alexandrak,D., Antoline,G., Anwar,R., Ballesta,J.P.,
REFERENCE  Benit,P. et al.
AUTHORS  The complete DNA sequence of yeast chromosome III
JOURNAL  Nature 357 (6373), 38-46 (1992)
MEDLINE  92244356
REFERENCE  9 (bases 1 to 316613)
AUTHORS  Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
TITLE    The complete sequence of a 9,543 bp segment on the left arm of
JOURNAL  chromosome III reveals five open reading frames including
MEDLINE  glucokinase and the protein disulfide isomerase
REFERENCE  yeast 8 (7), 569-575 (1992)
AUTHORS  92397595
MEDLINE  10 (bases 1 to 316613)
REFERENCE  Wilson,C., Grisanti,P. and Frontali,L.
AUTHORS  The complete sequence of a 6146 bp fragment of Saccharomyces
TITLE    cerevisiae chromosome III contains two new open reading frames
JOURNAL  yeast 8 (7), 569-575 (1992)
MEDLINE  92397594
REFERENCE  11 (bases 26740 to 32076)
AUTHORS  Defour,E., Debrabandere,R., Keyers,B., Voet,M. and Volckaert,G.
TITLE    Nucleotide sequence of D10B, a BamHI fragment on the small-ring
JOURNAL  chromosome III of Saccharomyces cerevisiae
MEDLINE  yeast 8 (8), 681-687 (1992)
REFERENCE  93070606
AUTHORS  12 (bases 169581 to 171116; 171683 to 172169)
TITLE    Agostoni Carbone,M.L., Panzeri,L., Muzi Falconi,M., Carcano,C.,
JOURNAL  Plevani,P. and Lucchini,G.
MEDLINE  Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III
REFERENCE  from strain AB972: evidence for a TY insertion and functional
AUTHORS  analysis of open reading frame YCR28
JOURNAL  yeast 8 (9), 805-812 (1992)
MEDLINE  93070619
REFERENCE  13 (bases 1 to 315338)
AUTHORS  Valle,G.
TITLE    TA-repeat microsatellites are closely associated with ARS consensus
JOURNAL  sequences in yeast chromosome III
MEDLINE  yeast 9 (7), 753-759 (1993)
REFERENCE  93377412
AUTHORS  14 (bases 1 to 315338)
TITLE    Slonimski,P.P. and Brouillet,S.
JOURNAL  A data-base of chromosome III of Saccharomyces cerevisiae
MEDLINE  yeast 9 (9), 941-1029 (1993)
REFERENCE  94091061
AUTHORS  15 (bases 1 to 315339)
TITLE    Rodriguez-Cousino,N., Lill,R., Neupert,W. and Court,D.A.
JOURNAL  Identification and initial characterization of the cytosolic
MEDLINE  protein Ycr77P
REFERENCE  yeast 11 (6), 581-585 (1995)
AUTHORS  95373282
TITLE    16 (bases 1 to 316613)
JOURNAL  MIPS.
MEDLINE  Direct Submission
REFERENCE  Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data
AUTHORS  collected by MIPS on behalf of the European Yeast Chromosome III
TITLE    Sequencing project
JOURNAL  replaced by [13]
MEDLINE  17 (bases 1 to 314957)
REFERENCE

```

	AUTHORS	Jimenez, A.
	TITLE	Direct Submission
	JOURNAL	Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from A. Jimenez
	REMARK	replaced by [14]
	REFERENCE	18 (bases 1 to 316613)
	AUTHORS	Louis, E.J.
	TITLE	Direct Submission
	JOURNAL	Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
	REMARK	revised by [18]
	REFERENCE	19 (bases 1 to 316613)
	AUTHORS	Louis, E.J.
	TITLE	Direct Submission
	JOURNAL	Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
	REMARK	Revised by [20]
	REFERENCE	20 (bases 1 to 316613)
	AUTHORS	Gromadka, R.
	TITLE	Direct Submission
	JOURNAL	Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis, Institute of Biochemistry and Biophysics, Pawinskiego 5a, Warsaw, Poland, Electronic Mail Address: robertepsd.ibb.waw.pl
	REMARK	revised by [21]
	REFERENCE	21 (bases 1 to 316613)
	AUTHORS	MIPS.
	TITLE	Direct Submission
	JOURNAL	Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF - Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany, mips-yeast-adm@gsf.de
	REMARK	Resequencing project, achieved by the joint effort of G. Valles and G. Voickerts laboratories. Munich information center for protein sequences.
	COMMENT	On Jul 3, 2001 this sequence version replaced gi:1907116. The contig sequence which has been released to the file server in 1992 has been subject to a resequencing project, achieved by the joint effort of G. Valles and G. Voickerts laboratories. The following lines present a summary of the altered entites. The resequenced chrIII contig is 316613 bp with GCG-check: 429 GENETIC ENTITIES valid no longer: YCL012w (YCL014w + YCL012w -> YCL014w / C_B1367 + C_C231->C_A1536), PEP y, ORF in identical form now C-terminal part of YCL014w. (A frame shift in the nt sequence of YCL014w leads to an elongated polypeptide.) This elongated polypeptide encoded by the current YCL014w resembles the Bud3p polypeptide from the EMBL-entry (Accession:SC175801.ID:UI7580, total length 1636 aa) YCL006c, PEP y, ORF obsolete, is not an ORF in current Chromosome III contig. YCR062zw (YCR061w + YCR062w -> YCR061w / C_A583 + C_B120 -> C_B631), PEP y, ORF in identical form now C-terminal part of YCR061w. YCR068w-a (YCR068w + YCR068w-a -> YCR068w / C_C429 + C_A145 -> C_A520). PEP y, ORF in identical form now C-terminal part of YCR068w. YCL010c (C_P111, former ORF in region corresponding to coord. 307253-307387(C)) is obsolete, open reading frame is destroyed by single nt exchange resulting in a stop codon after 45 aa. NEW GENETIC ENTITIES: YCL26c-b (C_E193(YCL027c-a)), PEP y, new ORF YCL021w-a (C_C125). PEP y, new or f ALTERED GENETIC ENTITIES: YCL076w, PEP y, peptide elongated N-terminally by 51 aa YCL074w, PEP y, 3 aa exchanged YCL073c, PEP y, 4 aa exchanged YCL066c (C_F190->C_E260), PEP y, peptide elongated N-terminally by 70 aa YCL065w, PEP y, 1 aa exchanged YCL064c, PEP y, 3 aa exchanged YCL061c (C_DB53->C_E1096), PEP y, peptide elongated N-terminally by 243 aa YCL051w, PEP y, 11 aa altered, 6 aa exchanged, 1 aa additional, 4 aa less at the C-terminus

Query Match	6.3%	Score 132.4:	DB 8:	Length 316613:
Best Local Similarity	52.3%	Pred. No. 9e-26:		
Matches 342:	Conservative 0:	Mismatches 306:	Indels 6:	Gaps 2:
Y 1281	GGTCCTCTGCGCATCTCGCCCTTTCAGGTTTGGATATGACATGAAAGGCTGCTAGTCT	1340		
D 145589	GATTACACCTGGTATTTGGACATATCCACTTTGGTCTCGTAATGATGTGGGCAATCTGA	145530		
Y 1341	AGAAAGATCAGCTCATCTGCTCATGAAACAGGTCAGACTTTCATTAACAATTTGGAGAG	1400		
D 145529	AGACGAGATGATCAACCTTTATTAAGATATGGAAGTAACTAGTATGTGTAAGCTCTACTAGAAC	145470		
Y 1401	TGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAAGGGGAAAAGTT	1460		
D 145469	AGATACACAAAGAAATTTACCATGTGGGGAACAGGAGATCTAATACAACTAGCTCATGATTT	145410		
Y 1461	GGGTTTCTATCAGACTTTTGGTCCAAAGCAAGGATATCACACTTGGGGGATTTATGGCTTT	1520		
D 145409	GAATATGATGACAGATTTGGGACGAGGTCCTCAATTAACCATTACCTGGGGCTGTGGTTCTCT	145350		
Y 1521	GTCAGATATACCAATTTGAAATTCGACATCACCCTCTCCGTCAGACAGAGGGGAGAT	1580		
D 145349	TTCTAAATTTCCCTATTCGTAATATTTACGCATCATTTATTTGCCCTCCAGTTGGGGAATC	145290		
Y 1581	CGCACACGACCATCACATTGACCGTTAACAATTTTCC---GGGCAAGCTGGTGGATTTTGTGCT	1637		
D 145289	TGGCGCACACCATTCATTCGCGACCTCTCAAAACGTACATGACACTCTTGTTGACGTCTTTGT	145230		
Y 1638	GACACACTTTGGGAACCCAGAAAGATGACCTGCACAGGAACCTGCAGGCTATTTGCTTTTC	1697		
D 145229	ATTCCATAGTGGACAAAGAGAGATGAAGGATGAAGACACGTCAAAAGATTAACATGAC	145170		
Y 1698	AAAACCTACGAAAATAGTACTCTTAATCAAGTGATATTTTGGGATATTAACCTTCAGCACC	1757		
D 145139	TAACTTATGGGCAATACGACTCGCCGACGCTATTTTATTAAGTTACTTAAGTTGTTGATCC	145110		
Y 1758	TGGCTCCAGAGATT--ATTACAGCTCAGTACACATGGCAATGTGAAGGATATCGACAG	1814		
D 145109	AGGTGAAGGCAACCTACATTAAGTACGTTAAGTGAACATCCGGAATGCAACGACATTTGATCC	145050		
Y 1815	CACGTGATATGACAGATGCTGTGAATTAATTTATGTATCGAGGGCTGATCAGTTGGGTTA	1874		
D 145049	TTTCGACACATGATTAATGATGCTGAGATAATCTTGTATAGGGGCTTGAGAAAGAGAA	144990		
Y 1875	TGCAGATATTCCTCCATGCTGAAACCTAGTATCACAATTCATGATGGGCAAAATTT	1928		
D 144989	TGCTAGAGTTTGCAGAGAGACGATTAACCGATACGAGACTCAAGTTGGTAAGTT	144936		
RESULT 7				
LOCUS	D89130	1234 bp	mRNA	linear
DEFINITION	Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0503.			
ACCESSION	D89130			
VERSION	D89130.1	GI:1749467		
KEYWORDS	Schizosaccharomyces pombe (strain:PR45) cDNA to mRNA, clone_lib:library of H. Nojima clone:SY 0503.			
ORGANISM	Schizosaccharomyces pombe			



sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
Cosmid c688 is overlapped at the 3' end by cosmid c3G9, EMBL entry SPAC3G9, accession number AL021046.

## FEATURES

## source

```

1. .34034
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/map="IL"
/clone="cosmid c688"
1. .1280
/gene="SPAC688.01"
1. .1280
/partial
/gene="SPAC688.01"
/note="SPAC688.01, len:425, SIMILARITY: Saccharomyces cerevisiae, YCQ7_YEAST, hypothetical 107.9 kd protein.", (953 aa), fasta scores: opt: 1034, E():0, (46.5% identity in 430 aa)"
/codon_start=3
/label="SPAC688.01"
/product="hypothetical protein similar to yeast YCRO17C protein involved in sensitivity to certain drugs"
/protein_id="CAB90768.1"
/db_xref="gi:7768477"
/translation="SYPMVNHGMSYGLGIFTSYVLIYSPASIRISSPYSAKVMG GAFVITYLSLAHVAVVAYEPPGPIILRERSTYLLIFGNMLALVPAYSESEKPN KADSVVDIKQDSYRKRSKSLTGTCLMALKFRQIMPPYDTPYHKNKLF TAGIWTIHGDNEMFVASENRIRDAVRMDLDFGLSDTORLIMGRDLYLAHD LGMYADYGPDPKHTMGALLSKPEIYVNSHTLPSPOGELAPAIHATLDYGEIIV VYSHGOYSEOLDRLQSTELARIMRESPRLVFLGYVSNVGOEPORTLIFDTGMLD IEPAYDRMCOYIEFVGKRICGICARLHRSITDTDELQGTGLVTDLGRNRIDKEHV PESHRYPLEFSTGVNGHYHNNLVVHEPWYD"
/complement(join(1545. .1979, 2109. .2256, 2343. .2392))
/gene="SPAC688.02c"
/complement(join(1545. .1979, 2109. .2256, 2343. .2392))
/gene="SPAC688.02c"
/note="SPAC688.02c, len:210, SIMILARITY: Saccharomyces cerevisiae, Q12143, chromosome xvi reading frame ori ypl233w., (216 aa), fasta scores: opt: 183, E():9.1e-05, (22.7% identity in 185 aa)"
/codon_start=1
/label="SPAC688.02c"
/product="hypothetical protein with very low similarity to yeast ypl233w"
/protein_id="CAB90769.1"
/db_xref="gi:7768478"
/translation="MSSHGDASVLPKRIALESINDYSVKSVFQAIOEKLAIHLEPQA SREAGMGEDDOLIRKVELMNLIELSTSIIOENTYINGDANSLODPKNSDOIEPFL DIALRTVROOLIFNEVEDAHVAVRKRSPQVQYKAYVADAMCOOAFPLRNKVDYVSL ODKYENDEQSTSTVFRKEDIDREQTIAKAVYAKKMLPRVAKLEKTP"
/complement(1980. .1990)
/gene="SPAC688.02c"
/note="SPAC688.02c"
/complement(2103. .2108)
/note="ttaacataag, splice branch and acceptor"
/gene="SPAC688.02c"
/note="gtgagt, splice donor sequence"
/complement(2257. .2271)
/gene="SPAC688.02c"
/note="ttaacactataag, splice branch and acceptor"
/complement(2337. .2342)
/gene="SPAC688.02c"
/note="gtgtgt, splice donor sequence"
/complement(join(2596. .2953, 3083. .3297, 3381. .3422))
/gene="SPAC688.03c"
/complement(join(2596. .2953, 3083. .3297, 3381. .3422))
/note="SPAC688.03c, len:204, SIMILARITY: Homo sapiens, Q91400, amecr1 protein., (333 aa), fasta scores: opt: 443, E():3.4e-25, (40.3% identity in 186 aa)"

```

```

/codon_start=1
/label="SPAC688.03c"
/product="conserved hypothetical protein"
/protein_id="CAB90770.1"
/db_xref="gi:7768479"
/translation="MICWPSLKNESMNDKEVCYCCEVVAATLEHRKVRDYNKASWT RSIPLFVFSAGKHDKQKRGICGFRAPLPVNTLYTGRSKQAQFDEFRPISIGELA LLECOIDLVDPEIDPDLMEVGIGVSIKFTANGYSSYSLPSVAAEPRMQDEET LESLHKAGYGSIRSLQITATRYKSLGCGYEYELHNLGLG"
/complement(join(2662. .2953, 3083. .3273))
/gene="SPAC688.03c"
/note="SPAC688.03c"
/note="Match to PF01871 DUF51, Protein of unknown function score 49.41"
/complement(2954. .2976)
/gene="SPAC688.03c"
/note="ttaacacactttaaaataag, splice branch and acceptor"
/complement(3077. .3082)
/gene="SPAC688.03c"
/note="gtagt, splice donor sequence"
/complement(3298. .3315)
/gene="SPAC688.03c"
/note="ttaataataatgacaaag, splice branch and acceptor"
/complement(3375. .3380)
/gene="SPAC688.03c"
/note="gtatga, splice donor sequence"
/complement(4047. .4775)
/gene="SPAC688.04c"
/complement(4047. .4775)
/gene="SPAC688.04c"
/note="SPAC688.04c, len:242, SIMILARITY: Saccharomyces cerevisiae, Y1V8_YEAST, Glutathione transferase G711, (234 aa), fasta scores: opt: 274, E():3.5e-11, (27.0% identity in 222 aa)"
/codon_start=1
/label="SPAC688.04c"
/product="putative glutathione S transferase"
/protein_id="CAB90771.1"
/db_xref="gi:7768480"
/translation="MIVLHHLKNSRSTRIVMLLEELKVPYKIVYDRVGRAPAYTK LSPDGKSPVVDGCVYTESAILEHLVYRKVGSPKSEEDVALEKEIEMHNSSEAS LMPFTWASHVIDLSYNTPIFRIVRQVNGISKYSKFTFLADTIDHNLASNEF FAGQFTYADQMCFFIFEAQGRDYLSQPKYRNKRMRVSDPRACRIAARVBDNLT TLFSQVRYSHRPPTPEQVRSDE"
/complement(4164. .4775)
/gene="SPAC688.04c"
/note="match to PF00043 GST, Glutathione S-transferases. score 85.13"
/complement(join(5637. .6842, 6927. .6980))
/gene="SPAC688.06c"
/complement(join(5637. .6842, 6927. .6980))
/gene="SPAC688.06c"
/note="SPAC688.06c, len:419, predicted to contain coiled-coil region, SIMILARITY: S. cerevisiae, LOW to YB61_YEAST, (324 aa), fasta scores: opt: 125, E():0.76, (21.3% identity in 300 aa)"
/codon_start=1
/label="SPAC688.06c"
/product="hypothetical protein very putative Actin-related protein, regulator of microtubule stability"
/protein_id="CAB90772.1"
/db_xref="gi:7768481"
/translation="MSAEYIEVSSSPDIFDDDDMITIEPELKNPKDCKSKRRSY TECCERILRTSKDPESTQOQLVHNNCTPKYKHENLNAVDEEDPTENPILSFSD NESDLEPDLTRVAEDNDVLISRTSKNSASCRNPFESHAHSNREISSSCFTYH RKPOLFEKSLKLGKNSLEANSRSLIKELCSANSTENVCSTVDEIQQDRHSAGH SIDSTGOSNFSLEGDSATHKRRKTDNIEFSCSEFENRSLINLYAGIMDNRKNADNE AKSLKEKLENPEVEKLRIASVYGFKSSDSKATLIKIVESCLDAIDRSOSKKGKET PHDYLTISTKTVLEFDDIIVQOTHRALISVQVQADNSWIKIILYLSAIDVEPEQLMLK RKNLNVSLDLKSNCKDKYGVLMKSMW"
/complement(6843. .6857)
/gene="SPAC688.06c"

```

misc\_feature

```
misc_feature /note="ttaaactgactcttag, splice branch and acceptor"
              complement(6921..6926)
              /gene="SPAC688.06c"
              /note="gattga, splice donor sequence"
              complement(7359..10475)
```

```
complement(7359. 10475
/ene="SPAC688.06c"
/ote="gtatga, splice c
complement(7359. 10475
```





\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1
7957: contig of 7956 bp in length
7957: gap of unknown length
8057: contig of 5847 bp in length
13904: gap of unknown length
14003: contig of 3955 bp in length
17958: gap of unknown length
18058: contig of 4296 bp in length
18059: gap of unknown length
22355: gap of unknown length
22455: contig of 4135 bp in length
26589: gap of unknown length
26590: contig of 4559 bp in length
31249: gap of unknown length
31348: contig of 5208 bp in length
36556: gap of unknown length
36557: contig of 3727 bp in length
40383: gap of unknown length
40384: contig of 3932 bp in length
44415: gap of unknown length
44515: contig of 4523 bp in length
49038: gap of unknown length
49138: contig of 3211 bp in length
52349: gap of unknown length
52449: contig of 3727 bp in length
55176: gap of unknown length
55177: contig of 3272 bp in length
56276: gap of unknown length
60332: contig of 4056 bp in length
60333: gap of unknown length
62803: contig of 2371 bp in length
62903: gap of unknown length
62904: contig of 3718 bp in length
66721: gap of unknown length
66722: contig of 2299 bp in length
69120: gap of unknown length
69121: contig of 3100 bp in length
72220: gap of unknown length
72221: contig of 3287 bp in length
75607: gap of unknown length
75608: contig of 3253 bp in length
75708: gap of unknown length
78960: contig of 2720 bp in length
79060: gap of unknown length
81780: contig of 3260 bp in length
81781: gap of unknown length
85140: contig of 3260 bp in length
85141: gap of unknown length
85241: contig of 2643 bp in length
87983: gap of unknown length
87984: contig of 2779 bp in length
90762: gap of unknown length
90763: contig of 2224 bp in length
90863: gap of unknown length
93086: contig of 2844 bp in length
93186: gap of unknown length
96030: contig of 1940 bp in length
96130: gap of unknown length
98070: contig of 2740 bp in length
98170: gap of unknown length
98171: contig of 1958 bp in length
100910: gap of unknown length
101010: contig of 1680 bp in length
101011: gap of unknown length
102968: contig of 1458 bp in length
103069: gap of unknown length
104748: contig of 1458 bp in length
104749: gap of unknown length
104849: contig of 1996 bp in length
106306: gap of unknown length
106406: contig of 1732 bp in length
108402: gap of unknown length
108403: contig of 1741 bp in length
108502: gap of unknown length
110235: contig of 1741 bp in length
110334: gap of unknown length
112075: contig of 1433 bp in length
112076: gap of unknown length
112175: contig of 1433 bp in length
112176: gap of unknown length

```

```

* 113609 113708: gap of unknown length
* 113709 115299: contig of 1591 bp in length
* 115300 115399: gap of unknown length
* 115400 116670: contig of 1271 bp in length
* 116671 116770: gap of unknown length
* 116771 119076: contig of 2306 bp in length
* 119077 119176: gap of unknown length
* 119177 120752: contig of 1576 bp in length
* 120753 120852: gap of unknown length
* 120853 122336: contig of 1484 bp in length
* 122337 122436: gap of unknown length
* 122437 125169: contig of 2733 bp in length
* 125170 125269: gap of unknown length
* 125270 126815: contig of 1546 bp in length
* 126816 126915: gap of unknown length
* 126916 128187: contig of 1272 bp in length
* 128188 128287: gap of unknown length
* 128288 129839: contig of 1552 bp in length
* 129840 129939: gap of unknown length
* 129940 131660: contig of 1721 bp in length
* 131661 131760: gap of unknown length
* 131761 133894: contig of 2134 bp in length
* 133895 133994: gap of unknown length
* 133995 135658: contig of 1664 bp in length
* 135659 135758: gap of unknown length
* 135759 136862: contig of 1104 bp in length
* 136863 136962: gap of unknown length
* 136963 138466: contig of 1404 bp in length
* 138467 138604: gap of unknown length
* 138605 139704: contig of 1138 bp in length
* 139705 139704: gap of unknown length
* 139706 141083: contig of 1379 bp in length
* 141084 141183: gap of unknown length
* 141184 142432: contig of 1249 bp in length
* 142433 142532: gap of unknown length
* 142533 144347: contig of 1815 bp in length
* 144348 144447: gap of unknown length
* 144448 145945: contig of 1498 bp in length
* 145946 146045: gap of unknown length
* 146046 147800: contig of 1755 bp in length
* 147801 147900: gap of unknown length
* 147901 149217: contig of 1317 bp in length
* 149218 149317: gap of unknown length
* 149318 150337: contig of 1020 bp in length
* 150338 150437: gap of unknown length
* 150438 151919: contig of 1482 bp in length

```

Query Match 3.9%; Score 81.6; DB 2; Length 165893;  
 Best Local Similarity 78.2%; Pred. No. 1.7e-11;  
 Matches 111; Conservative 0; Mismatches 29; Indels 2; Gaps 1;

```

QY 1882 ATCTCCATGCGACGACGATGATTCAGAAATTCAGATGCGCAAAATTTATCCCTGAT 1941
DB 165892 ATTTCATGCTGACGACGATGATCTGATGAT--CAGATGCCAAATTCAGATCCAGAT 165835
QY 1942 GACCCCAATTAATATAGACACCAAGAGTGTATAGACACAGAGAACTTCTGAG 2001
DB 165834 GATCCCTCAATTAACAGACCAAGAGTGTATAGACACAGAGATTTCCCAAG 165775
QY 2002 AAAATTCATTTTAATCCAGAT 2023
DB 165774 AACATTCATTTCAACCCGAGT 165753

```

```

RESULT 12
HMMTPPSA/c HMMTPPSA 1348 bp DNA linear PRI 14-JAN-1995
LOCUS DEFINITION Human TPI (triosephosphate isomerase) pseudogene psi-13C.
ACCESSION K03223
VERSION K03223.1 GI:339844
KEYWORDS pseudogene; triose-phosphate isomerase.
SOURCE Human DNA, clone psi-hrPI-13C.
ORGANISM Homo sapiens

```

REFERENCE 1 (bases 1 to 1348)  
AUTHORS Brown,J.R., Daar,I.O., Krug,J.R. and Maquat,L.E.  
TITLE Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family  
JOURNAL Mol. Cell. Biol. 5 (7), 1694-1706 (1985)  
MEDLINE 85267686

COMMENT The intronless pseudogene presented below, is missing 317 bases between positions 725 and 726, eliminating the stop codon. The start codon is located at positions 264-266. This gene is flanked by short direct repeats, 'taaatc', suggesting that it is generated by integration of a reverse-transcribed copy of TPI mRNA into germ line DNA via a transposon-like mechanism. A potential poly-A signal is located at positions 1119-1124.

FEATURES  
source 1..1348  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="12p13"

misc-feature 1  
/gene="TPI1"  
/note="G00-119-617"

COUNT 385 a 313 C 304 G 346 t  
ORIGIN 262 bp upstream of NcoI site.

Query Match 3.9%; Score 81.4; DB 9; Length 1348;  
Best Local Similarity 98.8%; Pred. No. 1.2e-11;  
Matches 82; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1187 TTCTGTGCTGCTTGTGGTGGGATTTGGATTAGACGACATTAAGCCTATG 1246  
|||||  
Db 169 TTCTGTGCTGCTTGTGGTGGGATTTGGATTAGACGACATTAAGCCTATG 110

QY 1247 AGAGAAACTGGGCAAGTGCA 1269  
|||||  
Db 109 AGAGAAACTGGGCAAGTGCA 87

RESULT 13  
AX106346/c 54 bp DNA linear PAT 30-APR-2001  
LOCUS AX106346  
DEFINITION Sequence 127 from Patent WO0125272.  
ACCESSION AX106346  
VERSION AX106346.1 GI:13922028  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 54)  
AUTHORS Xu,J., Skelky,Y.A., Reed,S.G. and Cheever,M.A.  
TITLE Compositions and methods for therapy and diagnosis of prostate cancer  
JOURNAL Patent: WO 0125272-A 127 12-APR-2001.  
CORIXA CORPORATION (US)

FEATURES  
source 1..54  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 23 a 17 c 9 g 5 t  
ORIGIN

Query Match 2.6%; Score 54; DB 6; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.00047;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 777  
|||||  
Db 54 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 1

RESULT 14

AX140637/c 54 bp DNA linear PAT 31-MAY-2001  
LOCUS AX140637  
DEFINITION Sequence 127 from Patent WO0134802.  
ACCESSION AX140637  
VERSION AX140637.1 GI:14280751  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 54)  
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,  
Skelky,Y.A. and Wang,A.

TITLE Compositions and methods for the therapy and diagnosis of prostate cancer  
JOURNAL Patent: WO 0134802-A 127 17-MAY-2001;  
CORIXA CORPORATION (US)

FEATURES  
source 1..54  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 23 a 17 c 9 g 5 t  
ORIGIN

Query Match 2.6%; Score 54; DB 6; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.00047;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 777  
|||||  
Db 54 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 1

RESULT 15  
AX200497/c 54 bp DNA linear PAT 29-AUG-2001  
LOCUS AX200497  
DEFINITION Sequence 127 from Patent WO0151633.  
ACCESSION AX200497  
VERSION AX200497.1 GI:15390309  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 54)  
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,  
Stolk,J.A., Skelky,Y.A., Wang,A. and Meagher,M.J.

TITLE Compositions and methods for the therapy and diagnosis of prostate cancer  
JOURNAL Patent: WO 0151633-A 127 19-JUL-2001;  
CORIXA CORPORATION (US)

FEATURES  
source 1..54  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 23 a 17 c 9 g 5 t  
ORIGIN

Query Match 2.6%; Score 54; DB 6; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.00047;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 777  
|||||  
Db 54 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTTCGTCGT 1

Search completed: October 21, 2002, 20:20:04  
Job time : 4661 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 17:33:38 : Search time 342 Seconds  
(without alignments)  
10557.518 Million cell updates/sec

Title: US-09-809-638-1  
Perfect score: 2103  
Sequence: 1 atgacccgcgtgtgagaga.....ccaaactattatgaac 2103

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Archived: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N.GeneSeq\_032802.\*  
1: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384.8	18.3	635	23	AAS66555
2	142.8	6.8	670	23	AAS66554
3	54	2.6	54	19	AAV61215
4	54	2.6	54	19	AAV58600
5	54	2.6	54	21	AAA06363
6	54	2.6	54	22	AAS63571
7	54	2.6	54	22	AAS10122
8	54	2.6	54	22	AAH93479
9	54	2.6	54	22	AAH84793

C	10	54	2.6	54	22	AAH02544
C	11	38.4	1.8	7850	23	ABL24162
C	12	38	1.8	2523	23	AAS72985
C	13	38	1.8	5813	20	AAI33059
C	14	37.8	1.8	6533	24	ABL33708
C	15	37.2	1.8	616	21	AAZ80484
C	16	37.2	1.8	6583	22	ABA19235
C	17	37	1.8	236303	22	AAS11614
C	18	36.6	1.7	6668	22	AAS46417
C	19	36.6	1.7	6668	24	ABL33218
C	20	36.4	1.7	32179	22	ABA17139
C	21	36.4	1.7	80331	22	AAC89559
C	22	36.2	1.7	8155	18	AAV74374
C	23	36	1.7	417	22	AAS27442
C	24	36	1.7	1734	22	AAH14130
C	25	36	1.7	2805	21	AAC81736
C	26	36	1.7	4236	20	AAS46427
C	27	36	1.7	6018	22	AAH84253
C	28	36	1.7	14041	22	AAH48024
C	29	35.8	1.7	1812	23	AAS67909
C	30	35.8	1.7	5263	22	AAS46468
C	31	35.8	1.7	5263	24	AAS61226
C	32	35.6	1.7	16283	24	AAS61363
C	33	35.6	1.7	56506	21	AAH69168
C	34	35.6	1.7	151826	21	AAF22291
C	35	35.4	1.7	5276	24	ABL33150
C	36	35.4	1.7	9300	22	ABA15594
C	37	35.4	1.7	127197	22	AAI61370
C	38	35.2	1.7	296	23	AAD23551
C	39	35.2	1.7	846	23	AAS76980
C	40	35.2	1.7	4015	22	AAH33952
C	41	35.2	1.7	4029	22	AAI58323
C	42	35.2	1.7	4331	22	AAI60109
C	43	35.2	1.7	6197	22	ABL33710
C	44	35	1.7	415	24	AAK76562
C	45	35	1.7	441	22	AAK76563

## ALIGNMENTS

RESULT 1  
ID AAS66555  
AAS66555 standard: cDNA: 635 BP.  
AC AAS66555:  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
XX  
DE DNA encoding novel human diagnostic protein #2359.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
PN WO200175067-A2.  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001: 2001WO-US08631.  
PF  
XX  
XX 31-MAR-2000: 2000US-0540217.  
PR 23-AUG-2000: 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX  
XX P-PSDB: ABG02368.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

Claim 1: SEQ ID No 2359; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Seq Sequence 635 BP; 255 A; 107 C; 112 G; 160 T; 1 other;

Query Match 18.3%; Score 384.8; DB 23; Length 635;  
Best Local Similarity 98.8%; Pred. No. 1.6e-105;

Matches 408; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

OY 1693 GTTCAAACTACTGAAAGTAGCTTATCAAGTATTTTGGGATATATCACTTCA 1752

DB 1 GTTCAAACTACTGAAAGTAGCTTATCAAGTATTTTGGGATATATCACTTCA 60

OY 1753 GCACCTGGCTCAGAGATATCTACAGCTCAGTGAATGGAATGTAAGATATCGAC 1812

DB 6 GCACCTGGCTCAGAGATATCTACAGCTCAGTGAATGGAATGTAAGATATCGAC 120

OY 1813 AGCAGTATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1872

DB 121 AGCAGTATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

OY 1873 TATGCAAGAAATCT-CCCATGCTGAACTGATGATGATGATGATGATGATGATGATGAT 1931

DB 181 TATGCAAGAAATCT-CCCATGCTGAACTGATGATGATGATGATGATGATGATGATGAT 240

OY 1933 GATCCCTGATGACCCCACTAATATATAGACAAACGAAGTGTATGATGATGATGATGATGAT 1991

DB 241 GATCCCTGATGACCCCACTAATATATAGACAAACGAAGTGTATGATGATGATGATGATGAT 300

OY 1992 AGTTCTGAGAAATTCATTTTATCCAGATTTGGATCTCAAGAAAGACACATTA 2051

DB 301 AGTTCTGAGAAATTCATTTTATCCAGATTTGGATCTCAAGAAAGACACATTA 360

OY 2052 TGAANAACCACTCATTTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2103

DB 361 TGAANAACCACTCATTTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413

RESULT 2

AAS6554  
ID AAS6554 standard; cDNA; 670 BP.

XX AAS6554;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #2358.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Yang YT;

DR WPI; 2001-639362/73.

PT P-PSDB; AAG02367.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

Claim 1: SEQ ID No 2358; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers, and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Seq Sequence 670 BP; 181 A; 177 C; 132 G; 180 T; 0 other;

Query Match 6.8%; Score 142.8; DB 23; Length 670;

Best Local Similarity 85.9%; Pred. No. 2e-32; Matches 170; Conservative 0; Mismatches 27; Indels 1; Gaps 1;

OY 1821 TCATCAGAGATGCTGTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1880

DB 349 TCAGGTTATTAACCTGCTTTTAAATACGCTTGATGATGATGATGATGATGATGATGATGAT 408

OY 1881 AATCTCCAGTGTGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1940

DB 409 AATCTCCAGTGTGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 468

OY 1941 TGACCCCACTAATATATAGACAAACGAAGTGTATGATGATGATGATGATGATGATGATGAT 2000

DB 469 TGACCCCACTAATATATAGACAAACGAAGTGTATGATGATGATGATGATGATGATGATGAT 527

OY 2001 GAAATTCATTTTAACCC 2018

DB 528 GAAATTCATTTTAACCC 545

RESULT 3  
AAV61215/C  
ID AAV61215 standard; cDNA: 54 BP.  
XX  
AC AAV61215;  
XX  
DT 06-JAN-1999 (first entry)  
XX  
DE cDNA sequence of prostate tumour clone.  
XX  
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9837093-A2.  
XX  
PD 27-AUG-1998.  
XX  
PR 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI: 1998-609886/51.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
XX  
PT used in a vaccine for the treatment of prostate cancer  
XX  
XX  
XX Claim 3; Page 89; 130pp; English.  
XX  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 19; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 724 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTGTTGCTGTGT 777  
|||||  
Db 54 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTGTTGCTGTGT 1

RESULT 4  
AAV58600/C  
ID AAV58600 standard; cDNA: 54 BP.  
XX  
AC AAV58600;  
XX  
DT 08-DEC-1998 (first entry)  
XX  
DE Prostate tumour specific gene clone.  
XX  
KM Prostate tumour specific gene; human; prostate cancer; detection;  
KM therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9837418-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03690.

XX  
PR 09-FEB-1998; 98US-0904809.  
PR 25-FEB-1997; 97US-0806596.  
PR 01-AUG-1997; 97US-0904809.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI: 1998-480805/41.  
XX  
PT Novel human prostate specific tumour protein and fragments - useful  
XX for detecting and treating prostate cancers  
XX  
PS Claim 1; Page 95; 141pp; English.  
XX  
CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprising contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 19; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 724 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTGTTGCTGTGT 777  
|||||  
Db 54 CTGCTGTGCTGGCAAGTGATGATGCTTCATCTTGTGTGTTGCTGTGT 1

RESULT 5  
AAA06363/C  
ID AAA06363 standard; cDNA: 54 BP.  
XX  
AC AAA06363;  
XX  
DT 13-JUN-2000 (first entry)  
XX  
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:127.  
XX  
KM Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KM immunogenic; cytostatic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200004149-A2.  
XX  
PD 27-JAN-2000.  
XX  
PF 14-JUL-1999; 99WO-US15838.  
XX  
PR 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232180.  
PR 09-APR-1999; 99US-0288946.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;  
XX  
DR WPI: 2000-171268/15.  
XX  
PT New polypeptide useful for treating and diagnosing prostate cancer

PT comprises an immunogenic portion of prostate tumor protein -  
XX  
PS Claim 1; Page 143; 263pp; English.  
XX  
CC The present invention describes isolated polypeptides, comprising an  
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides  
CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express  
CC the polypeptides, antibodies against the polypeptides and vaccines  
CC comprising them can be used for inhibiting the development of prostate  
CC cancer in a patient. The polypeptides can be used to generate antibodies  
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of  
CC the polynucleotides encoding the polypeptides can be used as a probe or  
CC to modulate the expression of the polypeptides. AAA06241 to AAA06591 and  
CC AY82000 to AY82020 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;  
  
Query Match 2.6%; Score 54; DB 21; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 724 CTGCTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTGCTGCT 777  
DB 54: CTGCTGCTGGCAAGTGATGATGATGCTTCATCTTGTGTTGCTGCTGCT 1  
  
RESULT 6  
AAS63571/c  
ID AAS63571 standard; cDNA: 54 BP.  
XX  
AC AAS63571;  
XX  
DT 29-JAN-2002 (first entry)  
XX  
DE Human prostate cDNA sequence #123.  
XX  
KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200173032-A2.  
XX  
PD 04-OCT-2001.  
XX  
PF 27-MAR-2001; 2001WO-US09919.  
XX  
PR 27-MAR-2000; 2000US-0536857.  
PR 09-MAY-2000; 2000US-0568100.  
PR 12-MAY-2000; 2000US-0570737.  
PR 13-JUN-2000; 2000US-0593793.  
PR 27-JUN-2000; 2000US-0605783.  
PR 10-AUG-2000; 2000US-0636215.  
PR 29-AUG-2000; 2000US-0651236.  
PR 06-SEP-2000; 2000US-0657279.  
PR 02-OCT-2000; 2000US-0679426.  
PR 10-OCT-2000; 2000US-0685166.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick JS, Carter D;  
PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;  
XX  
DR WPI: 2001-639232/73.  
XX  
PT New human prostate-specific polypeptides and polynucleotides useful for  
XX the diagnosis and treatment of cancer, especially prostate cancer -  
XX  
PS Claim 1; Page 274; 579pp; English.

CC The invention relates to isolated prostate-specific  
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,  
CC antibodies raised against the polypeptides (or antigenic epitopes  
CC derived from them) and antigen-presenting cells expressing the  
CC polypeptides. The antibodies are useful for detecting the presence of  
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
CC the antigen-presenting cells are useful for stimulating and/or expanding  
CC T cells specific for a tumour protein, and for inhibiting the development  
CC of cancer especially prostate cancer. Compositions comprising the  
CC polynucleotide and/or polypeptide are useful for stimulating an immune  
CC response, and for treating cancer. The oligonucleotide is useful for  
CC detecting cancer. The present sequence is a prostate specific  
CC polynucleotide of the invention.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;  
  
Query Match 2.6%; Score 54; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 724 CTGCTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTGCTGCT 777  
DB 54: CTGCTGCTGGCAAGTGATGATGATGCTTCATCTTGTGTTGCTGCTGCT 1  
  
RESULT 7  
AAS10122/c  
ID AAS10122 standard; cDNA: 54 BP.  
XX  
AC AAS10122;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human prostate tumour cDNA #13.  
XX  
KW Human; prostate tumour protein; prostate cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN US6362245-B1.  
XX  
PD 17-JUL-2001.  
XX  
PF 25-FEB-1998; 98US-0030607.  
XX  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
PR 09-FEB-1998; 98US-0020956.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Xu J, Dillon DC;  
XX  
DR WPI: 2001-440862/47.  
XX  
PT Novel polynucleotide encoding polypeptide comprising a portion of  
PT prostate tumour protein useful for inhibiting development of prostate  
PT cancer or for treating prostate cancer in a patient -  
XX  
PS Example 2; Column 137; 105pp; English.  
XX  
PI The sequence is a human prostate tumour cDNA which encodes a  
CC partial tumour protein. The DNA is useful for inhibiting the development  
CC of prostate cancer or for treating prostate cancer in a patient.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;  
  
Query Match 2.6%; Score 54; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 724 CTGCTGCTGGCAAGTGATGATGCTTCATCTTGTGTTGCTGCTGCT 777  
|||||

DB 54 CTGCTGTGCTTGGCAAGTGATGATGCTTCACCTGTTGTGCTTGTGCTGT 1

RESULT 8  
AAH93479/C  
ID AAH93479 standard; CDNA: 54 BP.  
XX  
XX AAH93479;  
XX  
XX 04-OCT-2001 (first entry)  
XX  
XX Human prostate-specific cDNA sequence p126.  
XX  
XX Human; prostate cancer; prostate-specific; diagnosis; vaccine;  
XX  
XX cytostatic; gene therapy; metastasis; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200151633-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 16-JAN-2001; 2001WO-US01574.  
XX  
XX 14-JAN-2000; 2000US-0483672.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;  
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;  
PI Wang A, Meagher MJ;  
XX  
XX WPI: 2001-425873/45.  
XX  
XX New polynucleotide encoding a prostate-specific protein, for  
PT diagnosing, monitoring and treating prostate cancer in a patient and  
PT for use in vaccines -  
XX  
XX  
XX Claim 1; Page 272; 543pp; English.  
XX  
XX The present invention describes polynucleotide sequences (I) which encode  
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,  
CC and can be used in vaccine production and gene therapy. (I), (II),  
CC antibodies to (II), fusion proteins comprising (II), and isolated  
CC T cells prepared using (I) or (II) are used to treat cancer in a  
CC patient. The cancer that is diagnosed or treated is particularly  
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
CC (I) can be used for monitoring the progression of cancer in a patient.  
CC (I) and (II) can also be used to improve diagnostic and therapeutic  
CC methods for prostate cancer. They can indicate the level of metastasis  
CC as well as the prostate volume. AAH93357 to AAH93944 and AAH01115 to  
CC AAH01318 represent polynucleotide and amino acid sequences used in the  
CC exemplification of the present invention.  
XX  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 724 CTGCTGTGCTTGGCAAGTGATGATGCTTCACCTGTTGTGCTTGTGCTGT 777  
DB 54 CTGCTGTGCTTGGCAAGTGATGATGCTTCACCTGTTGTGCTTGTGCTGT 1

RESULT 9  
AAH84793/C  
ID AAH84793 standard; CDNA: 54 BP.  
XX  
XX AAH84793;  
XX  
XX 25-SEP-2001 (first entry)

XX  
XX Human prostate-specific cDNA sequence p126.  
DE  
XX  
XX Human; prostate cancer; therapy; diagnosis; cat eye syndrome;  
XX  
XX chromosome 22q11.2; prostate-specific protein; chromosome 1;  
XX  
XX prostate specific antigen; PSA; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200134802-A2.  
XX  
XX 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US30904.  
XX  
XX 12-NOV-1999; 99US-0439313.  
XX  
XX 18-NOV-1999; 99US-0443686.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;  
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;  
XX  
XX WPI: 2001-308785/32.  
XX  
XX Isolated polypeptide comprising at least an immunogenic portion of a  
PT prostate-specific protein, useful in the diagnosis and therapy of a  
PT prostate cancer -  
XX  
XX  
XX Claim 5; Page 172; 325pp; English.  
XX  
XX The present invention describes an isolated polypeptide (PI) comprising  
CC at least an immunogenic portion of a prostate-specific protein, or its  
CC variant. Also described are polynucleotides (NI) encoding (PI). (PI) and  
CC (NI) have cytostatic activity and can be used in vaccine production.  
CC The polypeptides, nucleic acids and antibodies from the present  
CC invention are useful in the diagnosis and therapy of prostate cancer.  
CC Prostate specific genes P704P, P712P, P774P, P775P and B305P are located  
CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome  
CC region. Prostate specific antigen (PSA) P501S was located on  
CC chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent  
CC polynucleotide and polypeptide sequences used in the exemplification  
CC of the present invention.  
XX  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 724 CTGCTGTGCTTGGCAAGTGATGATGCTTCACCTGTTGTGCTTGTGCTGT 777  
DB 54 CTGCTGTGCTTGGCAAGTGATGATGCTTCACCTGTTGTGCTTGTGCTGT 1

RESULT 10  
AAH02544/C  
ID AAH02544 standard; CDNA: 54 BP.  
XX  
XX AAH02544;  
XX  
XX 14-JUN-2001 (first entry)  
XX  
XX Prostate tumour antigen determined cDNA sequence for p126.  
XX  
XX Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;  
XX  
XX prostate cancer; immunogenic; cytostatic; vaccine; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200125272-A2.  
XX  
XX 12-APR-2001.

XX 04-OCT-2000; 2000MO-US27464.  
XX 04-OCT-1999; 99US-0157455.  
XX (CORI-) CORIXA CORP.  
XX Xu J, Skelky YW, Reed SG, Cheever MA:  
XX WPI: 2001-245062/25.  
XX  
XX Prostate specific protein and its encoding polynucleotide, useful for  
XX the treatment and diagnosis of prostate cancer -  
XX  
XX Claim 4; Page 162; 276pp: English.  
XX  
XX The present invention describes an isolated polypeptide (I) comprising  
XX at least an immunogenic portion of a prostate tumour antigen protein or  
XX its variant, (I) have cytostatic activity and can be used in vaccine  
XX production, (I), prostate tumour antigen polynucleotides, an antigen  
XX presenting cell (APC e.g. a dendritic cell) that expresses (I), and a  
XX pharmaceutical composition containing (I) are useful for inhibiting the  
XX development of cancer in a patient. Antibodies specific for prostate  
XX specific proteins and oligonucleotides that hybridise to a  
XX polynucleotide that encodes a prostate specific protein are useful  
XX for detecting the presence or absence of a cancer or monitoring the  
XX progression of a cancer, especially prostate cancer.  
XX AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences  
XX used in the exemplification of the present invention.  
XX  
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;  
Query Match 2.6%; Score 54; DB 22; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 724 CTGCTGCTGCTGGCAAGTGATGATGCTTCATCTGTTGTGGTTCGT 777  
DB 54 CTGCTGCTGCTGGCAAGTGATGATGCTTCATCTGTTGTGGTTCGT 1  
RESULT 11  
ABL24162  
ID ABL24162 standard; DNA; 7850 BP.  
XX  
XX ABL24162;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23959.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI: 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell

PT Interactions -  
XX  
XX Claim 1; SEQ ID NO 23959; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
XX sequences (AB101840-AB16175) and the encoded proteins  
XX (AB57737-AB572072).  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 7850 BP; 1994 A; 1770 C; 1831 G; 2255 T; 0 other;  
Query Match 1.8%; Score 38.4; DB 23; Length 7850;  
Best Local Similarity 50.0%; Pred. No. 2.8;  
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
OY 1070 TGTATTATATGCGGCTGAATATGCTATTGCTCTAAGAAAACCTTGACTGCTTTC 1129  
DB 5814 TGTATTATATGCGGCTGAATATGCTATTGCTCTAAGAAAACCTTGACTGCTTTC 1189  
OY 1130 AAACAAAACAGTCTTAAGTCTTTTCAGAAAGTGAATAATACATGAACCTTTTC 1189  
DB 5874 GTTTCGAATATGATTAAGCGGAGGTTTGAAGTAATATGTCATATCTTGCATATTACA 5933  
OY 1190 TGTGCTGCTGTTGTTGTTGGATTTGGGATTTAGACACTACGATTAAGCCTATGACA 1249  
DB 5934 CTTAATGCTGACACCTCTTGTGATTAAGTAATAGCCTTGTCTATGATGACTGCTA 5993  
OY 1250 GAAACGTGGCA 1261  
DB 5994 GAAATGTGTCA 6005  
RESULT 12  
AAS72985/C  
ID AAS72985 standard; cDNA; 2523 BP.  
XX  
XX AAS72985;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #8789.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX P-PSDB; ABG08798.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess



PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 8789; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 2523 BP; 526 A; 803 C; 669 G; 525 T; 0 other:  
Query Match 1.8%; Score 38; DB 23; Length 2523;  
Best Local Similarity 49.0%; Pred. No. 1.9;  
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
OY 640 GGAGAGCTCTCTCTTTCCAGATGGGACAGTGGGACATCCACAGGGCAGAT 699  
DB 598 GGAGAGCTCTCTCTCTTTCCAGATGGGACAGTGGGACATCCACAGGGCAGAT 539  
OY 700 CCTAACCCATTGAGAGTCACTGCTGCTGGCAAGTGATGATGCTTCANCT 759  
DB 538 GATGAGAGAGTGTGGCTTCAAGGGGCTAGTGTGCTGAGAGCTACAGAGCAAGAGCA 479  
OY 760 TGT 819  
DB 478 TGGCTGT 419  
OY 820 GGGCTCTTTTACCTGCACACATGGGC 845  
DB 418 GTGGCAGACATCTGCAGAAACGGGC 393  
RESULT 13  
AAI13059/C  
ID AAI13059 standard; DNA; 9813 BP.  
XX  
AC AAI13059;  
XX  
DT 19-MAR-1999 (first entry)  
XX  
DE Enterococcus faecalis genome contig SEQ ID NO:122.  
XX  
KW Enterococcus faecalis; config; detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
XX  
OS Enterococcus faecalis.  
XX  
PN WO9850555-A2.  
XX  
PD 12-NOV-1998.  
XX  
PF 04-MAY-1998; 98WO-US08985.  
XX  
PR 14-NOV-1997; 97US-0066009.  
PR 06-MAY-1997; 97US-0044031.  
PR 16-MAY-1997; 97US-0046555.

XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Dillon PJ, Kunsch CA;  
XX  
DR WPI: 1999-045171/04.  
XX  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT infection.  
XX  
PS Claim 1; Page 734-739; 2084pp; English.  
XX  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AAI12938 to AAI13919 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
XX  
SQ Sequence 9813 BP; 3233 A; 1698 C; 2161 G; 2719 T; 2 other:  
Query Match 1.8%; Score 38; DB 20; Length 9813;  
Best Local Similarity 49.0%; Pred. No. 4.1;  
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
OY 1106 AGAAAACCTTGACTTCTTCAACAAAACAGCTTCAAGTCTTTCAGAAAGA 1165  
DB 2747 AAAAATTCGAGATTTTTCATTAATGAAATAGTTCAGATTTCTTTAAAAAAA 2688  
OY 1166 GTGAAAATACATCAAACTTTTCTGTGCTCTGTTGGTGTGGATTTGGGATTAG 1225  
DB 2687 GAAATTAAGAGAAAACCTTGTCTCTTGTGTTCTTGTGATTTTGTGTTGCT 2628  
OY 1226 GACTACGGCATTAAGCTGTGAGAAACACTGGGCAAGTGGCACCACCAAGAGGCT 1285  
DB 2627 TGGACCATGACACACACTGGGATTAATTTGACTAATAGAAATTAACCAATAAATT 2568  
OY 1286 CTGCTGCCATCTGGCCCTTTCAGGTTT 1311  
DB 2567 AGGTAATTAATCTTGCACCTGCTGCTGT 2542  
RESULT 14  
ABL33708/C  
ID ABL33708 standard; DNA; 6533 BP.  
XX  
AC ABL33708;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1681.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianemic; cyostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX

PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI: 2002-130909/17.  
 XX  
 PR Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1681; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 SQ Sequence 6533 BP; 1675 A; 132 C; 1539 G; 3187 T; 0 other:  
 Query Match 1.8%; Score 37.8; DB 24; Length 6533;  
 Best Local Similarity 49.7%; Pred. No. 3.8;  
 Matches 96; Conservative 0; Mismatches 97; Indels 0; Gaps 0:  
 OY 1910 AATTTCAGTGGCAAAATTAGATCCCTGATGACCCCACTAATTATAGACAAACAGCA 1969  
 DB 6224 AAACACTACATATAAAATCTACTATCACCACCAACACGACATTTTAAACCGAAA 6165  
 OY 1970 AAGTGTCTATAGACCAAGAGAGTTTCGAGAAATTCATTTTAATCCAGATTGGAT 2029  
 DB 6164 AACAAACACATACCAATATAAATCCCAAAATATAAATCAAAACACAAAAA 6105  
 OY 2030 CCTACAAAGAGACACATATATGAAACACACATCATTTATGATAGATCCCAAT 2089  
 DB 6104 CCAACAAAAAATATACAAATATATATAAAACCAAAATTTCTTAACACAGCACTTAA 6045  
 OY 2090 ACTTTTATGAAA 2102  
 6044 AATTATTTTAATA 6032  
 RESULT 15  
 AA280484  
 ID AA280484 standard; cDNA; 616 BP.  
 XX  
 AC AA280484;  
 XX  
 DT 07-APR-2000 (first entry)  
 XX  
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:568.  
 XX  
 KW Human gene expression product; diagnosis; tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cyclostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9964576-A2.  
 XX  
 PD 16-DEC-1999.

XX  
 PF 09-JUN-1999; 99WO-IB01062.  
 XX  
 PR 10-JUN-1998; 98US-0088801.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
 PI Schlegel R,  
 XX  
 DR WPI: 2000-087220/07.  
 XX  
 PR Novel nucleic acids, used to develop products for the diagnosis and  
 PT treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer -  
 XX  
 PS Claim 15; Page 356; 469pp; English.  
 CC AA279917 to AA280766 represent double stranded cDNA clones isolated from  
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g.  
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
 CC can be used for developing agents for the diagnosis and treatment of  
 CC disorders involving unwanted cell proliferation, such as neoplasia,  
 CC dysplasia or hyperplasia.  
 CC  
 SQ Sequence 616 BP; 145 A; 142 C; 163 G; 144 T; 22 other:  
 Query Match 1.8%; Score 37.2; DB 21; Length 616;  
 Best Local Similarity 58.3%; Pred. No. 1.5;  
 Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0:  
 OY 1106 AGAAAAACCTGACCTGCTTCTTCAACAAAAAAGAGTTCTAAGTCTTTGACAAAGA 1165  
 DB 290 AGACAGATCTGCTTACCATCTTTAACAAGCAAAATGCTCTCTCTTCAGAGAAAGA 349  
 OY 1166 GTGAATAATACATGAACCTTTCTGTCGCTGCTGCTGGTGGAT 1213  
 DB 350 GTCAATTAACACTAAATCTTTCTTNGAAGTTCTCTTTCCGAT 397  
 Search completed: October 21, 2002, 19:03:07  
 Job time : 400 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:56:23 : Search time 69 seconds

(without alignments)  
7486,482 Million cell updates/sec

Title: US-09-809-638-1

Perfect score: 2103

Sequence: 1 atgacctcgtgtgtgagaga.....ccaaactttttatgaac 2103

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCBUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	2.6	54	US-09-020-956-127	Sequence 127, App
2	54	2.6	54	US-09-030-607-127	Sequence 127, App
3	54	2.6	54	US-09-439-313-127	Sequence 127, App
4	38.2	1.8	7218	US-08-237-463-14	Sequence 14, Appl
5	37.2	1.8	616	US-09-328-111-568	Sequence 568, App
6	37.2	1.8	7218	US-08-232-463-14	Sequence 14, Appl
7	34.8	1.7	1858	US-08-359-705B-7	Sequence 7, Appl
8	34.8	1.7	1858	US-08-286-846A-7	Sequence 7, Appl
9	34.8	1.7	1858	US-08-457-880A-7	Sequence 7, Appl
10	34.8	1.7	1858	US-08-444-622A-7	Sequence 7, Appl
11	34.8	1.7	1858	US-08-942-562-7	Sequence 7, Appl
12	34.8	1.7	1858	US-09-156-923-7	Sequence 7, Appl
13	34.8	1.7	2715	US-08-359-705B-5	Sequence 5, Appl
14	34.8	1.7	2715	US-08-286-846A-5	Sequence 5, Appl
15	34.8	1.7	2715	US-08-457-880A-5	Sequence 5, Appl
16	34.8	1.7	2715	US-08-444-622A-5	Sequence 5, Appl
17	34.8	1.7	2715	US-08-942-562-5	Sequence 5, Appl
18	34.8	1.7	2715	US-09-156-923-5	Sequence 5, Appl
19	34.8	1.7	2940	US-08-286-305A-8	Sequence 8, Appl
20	34.8	1.7	2940	US-08-441-104A-8	Sequence 8, Appl
21	34.8	1.7	2940	US-08-440-816A-8	Sequence 8, Appl
22	34.8	1.7	2940	US-09-417-381A-8	Sequence 8, Appl
23	33.6	1.6	1694	US-09-362-473-3	Sequence 3, Appl
24	33.6	1.6	2094	US-09-106-194-1	Sequence 1, Appl
25	32.8	1.6	1683	US-09-009-443-6	Sequence 6, Appl
26	32.8	1.5	2310	US-08-416-581B-2	Sequence 2, Appl
27	32	1.5	2787	US-08-416-581B-3	Sequence 3, Appl

28	32	1.5	2787	US-08-416-581B-4	Sequence 4, Appl
29	32	1.5	2787	US-09-288-461-1	Sequence 1, Appl
30	32	1.5	2787	US-09-087-465-5	Sequence 5, Appl
31	31.8	1.5	1480	US-08-484-105-19	Sequence 19, Appl
32	31.8	1.5	1480	US-08-484-106-19	Sequence 19, Appl
33	31.8	1.5	3437	US-08-860-339-17	Sequence 17, Appl
34	31.6	1.5	151	US-09-439-313-319	Sequence 319, App
35	31.4	1.5	533	US-09-020-956-71	Sequence 71, Appl
36	31.4	1.5	533	US-09-030-607-71	Sequence 71, Appl
37	31.4	1.5	533	US-09-439-313-71	Sequence 71, Appl
38	31	1.5	1667	US-08-485-284A-1	Sequence 1, Appl
39	30.8	1.5	580	US-09-228-986-23	Sequence 23, Appl
40	30.8	1.5	2690	US-09-029-755C-1	Sequence 1, Appl
41	30.8	1.5	3539	US-09-508-542-16	Sequence 16, Appl
42	30.8	1.5	19011	US-08-310-356-36	Sequence 36, Appl
43	30.8	1.5	19557	PCT-US92-06300-1	Sequence 1, Appl
44	30.6	1.5	1635	US-08-974-180-16	Sequence 16, Appl
45	30.6	1.5	2178	US-08-974-180-11	Sequence 11, Appl

## ALIGNMENTS

```
RESULT 1
US-09-020-956-127/c
; Sequence 127, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020, 956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-020-956-127

Query Match 2.6%; Score 54; DB 4; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.8e-07;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 724 CTGCTGCTGGCAAGTGGATGGATGCTTCATCTTGTGGTTCGTCGT 777
DB 54 CTGCTGCTGGCAAGTGGATGGATGCTTCATCTTGTGGTTCGTCGT 1
```

RESULT 2  
US-09-030-607-127/c  
Sequence 127, Application US/09030607  
Patent No. 6262245  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
NUMBER OF SEQUENCES: 224  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/030,607  
FILING DATE: 25-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.427C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-030-607-127  
Query Match 2.6%; Score 54; DB 4; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.8e-07;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 724 CTGCTGCTTGGCAAGTGATGATGCTTCATCTTGTGTGCTTCGTGT 777  
54 CTGCTGCTTGGCAAGTGATGATGCTTCATCTTGTGTGCTTCGTGT 1  
RESULT 3  
US-09-439-313-127/c  
Sequence 127, Application US/09439313  
Patent No. 6329505  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan Louise  
APPLICANT: Jiang Yugu  
APPLICANT: Reed, Steven G.  
APPLICANT: Kalos, Michael  
APPLICANT: Fanger, Gary  
APPLICANT: Retter, Mark  
APPLICANT: Solk, John  
APPLICANT: Day, Craig  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
FILE REFERENCE: 210121.427C9

CURRENT APPLICATION NUMBER: US/09/439,313  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 575  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 127  
LENGTH: 54  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-439-313-127  
Query Match 2.6%; Score 54; DB 4; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.8e-07;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 724 CTGCTGCTTGGCAAGTGATGATGCTTCATCTTGTGTGCTTCGTGT 777  
Db 54 CTGCTGCTTGGCAAGTGATGATGCTTCATCTTGTGTGCTTCGTGT 1  
RESULT 4  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-Fls  
US-08-232-463-14  
Query Match 1.8%; Score 38.2; DB 1; Length 7218;  
Best Local Similarity 10.0%; Pred. No. 0.34;  
Matches 46; Conservative 200; Mismatches 213; Indels 0; Gaps 0;  
OY 1136 CTTCAACAAAACAGTTCTTAAGTCTTTTCAGAAAGTGAAAATACATGAACCTT 1185

Db 1490 CATCACTGTAATACCTTATGCACTAGTAAAGATTAAGAAATTTGGTACRRR 1431  
Oy 1186 TTTCTGCGCTGTTGTTGGTGGATTTGGATTAGCACTAGCCATTAACCTAT 1245  
Db 1430 RRR 1371  
Oy 1246 GAGAAACCTGGGCAAGTGGCAACCAAGAGTCTGCGCATCGGCTTTC 1305  
Db 1370 RRR 1311  
Oy 1306 AGTTTGATAGCAATGAGGCTGTCTAGTAAAGATCAGCTACCTGCTCAT 1365  
Db 1310 RRR 1251  
Oy 1366 GAACAGTGCAGATTATCAACAATTTGGAGAGTATGCTTAAACCTATATGGG 1425  
Db 1250 RRR 1191  
Oy 1426 AACATGACTTAACCATGTGGCTAGGGAAAGTTGGTTCTATACAGACTTGTCCA 1485  
Db 1190 RRR 1131  
Oy 1486 AGCAACAGTATACACTTGGGGATTAATGCTTGTCAAGATACCAATTTGTAATCT 1545  
Db 1130 RRR 1071  
Oy 1546 GAGCATCACCTTCTTCCTCACCAGAGGCGCATGCA 1584  
Db 1070 RRRATCGCAAGCTCCTCGACCTGACCAAGCTCGCA 1032

RESULT 5  
US-09-328-111-568  
Sequence 568, Application US/09328111  
Patent No. 6262333  
GENERAL INFORMATION:  
APPLICANT: Endege, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.  
APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Dertl, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
TITLE OF INVENTION: PRODUCTS  
FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/328.111  
CURRENT FILING DATE: 1999-06-08  
EARLIER APPLICATION NUMBER: US 60/088,801  
EARLIER FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 568  
LENGTH: 616  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(616)  
OTHER INFORMATION: n = A/T/C or G  
US-09-328-111-568

Query Match 1.8%; Score 37.2; DB 4; Length 616;  
Best Local Similarity 58.3%; Pred. No. 0.18;  
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
Oy 1106 AGAAACCTGACTGCTTCTTCAACAAACAAACAGCTTAAAGTCTTTCAGAAAGA 1165

Db 290 AGACGACTGCTGTTTACATCTTTTAAACAAAGCAATGCTTCTTCTTCAGAAAGA 349  
Oy 1166 GTGAAAAATACATGAACCTTTTCTGCGCTGCTGTTGGTGGAT 1213  
Db 350 GTCATTAACTAAATAATCTTTCTTNNAGAGTTCTTCTTCGAT 397

RESULT 6  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHETTLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)83-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 1.8%; Score 37.2; DB 1; Length 7218;  
Best Local Similarity 4.6%; Pred. No. 0.68;  
Matches 18; Conservative 203; Mismatches 171; Indels 0; Gaps 0;  
Oy 583 TGGCTGCTGGCAGGGGCTCTTTGGTACCTGTGCTTCTCACCACCTGGGTTTGA 642  
Db 1040 TGGCTGCTGCTGAGGAGCTGCAATTTTTTTTTTTTTTTTTTTTTT 1099  
Oy 643 GAAGCTCTCTGTTTCCAGATGGCGAGTGGCATCCACATCCAGCCAGATCT 702  
Db 1100 YY 1159  
Oy 703 AACCATTTGAGGTGACGACTGCTGCTTGGCAAGTGATGATGCTTCACATCTGT 762  
Db 1160 YY 1219



CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457.880A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,622  
FILING DATE: 19-May-1995  
APPLICATION NUMBER: 08/286846  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0873P1C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1858 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-457-880A-7

Query Match 1.7%; Score 34.8; DB 2; Length 1858;  
Best Local Similarity 58.8%; Pred. No. 1.7;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 580 AACTGCTCTGCGAGGGCTGTTGGTAGCTGTCTCCACCCAGGCTTTT 639  
DB 1215 AAGGGCTCTTGAGAGAGTGGCATTGAGTGTGGCTGCCACAGTGGTTTT 1156  
OY 640 GGAGAGTCTCTGTTCCAGATGGCGAGTGAAGGCGAT 681  
DB 1155 GGCAATGAGGTATAGTTCGCAATGTTGTAGTGGTGGCTT 1114

US-08-444-622A-7/C  
Sequence 7, Application US/08444622A  
Patent No. 6025166

GENERAL INFORMATION:  
APPLICANT: Leonard G. Presta  
APPLICANT: David L. Shelton  
APPLICANT: Roman Urfier  
TITLE OF INVENTION: HUMAN TRK RECEPTORS AND NEUROTROPHIC FACTOR INHIBITORS  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,622A  
FILING DATE: 19-May-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/286846  
FILING DATE: 5  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0873P1C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1858 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-444-622A-7

Query Match 1.7%; Score 34.8; DB 3; Length 1858;  
Best Local Similarity 58.8%; Pred. No. 1.7;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 580 AACTGCTCTGCGAGGGCTGTTGGTAGCTGTCTCCACCCAGGCTTTT 639  
DB 1215 AAGGGCTCTTGAGAGAGTGGCATTGAGTGTGGCTGCCACAGTGGTTTT 1156  
OY 640 GGAGAGTCTCTGTTCCAGATGGCGAGTGAAGGCGAT 681  
DB 1155 GGCAATGAGGTATAGTTCGCAATGTTGTAGTGGTGGCTT 1114

RESULT 11  
US-08-942-562-7/C  
Sequence 7, Application US/08942562  
Patent No. 6027927

GENERAL INFORMATION:  
APPLICANT: Presta, Leonard G.  
APPLICANT: Shelton, David L.  
APPLICANT: Urfier, Roman  
TITLE OF INVENTION: Human trk Receptors and Neurotrophic  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,562  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/444,597  
FILING DATE: 19-May-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0873P1C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1858 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single

TOPOLOGY: Linear  
US-08-942-562-7

Query Match 1.7%; Score 34.8; DB 3; Length 1858;  
Best Local Similarity 58.8%; Pred. No. 1.7;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 580 AACTGCTCTGTCAGAGGCGCTGCTTTGGTAGCTGTCTTCACCCAGTGGGTTTT 639  
DB 1215 AAGGGCTCTTGAGAGAACTGGCATGTATGCTGCTGCGTCGCCAGTGGGTTTT 1156  
QY 640 GGAGAAGTCTCTCTTTGTTCCAGATGGCGAGTGGGCAT 681  
DB 1155 GGCAATGAGGTATAGTCCCATTTGTTGATGGGGGCTT 1114

RESULT 12  
US-09-156-923-7/c

Sequence 7, Application US/09156923  
Patent No. 6153189

GENERAL INFORMATION:  
APPLICANT: Presta, Leonard G.

APPLICANT: Shelton, David L.

TITLE OF INVENTION: Human trk Receptors and Neurotrophic Factor Inhibitors

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California

COUNTRY: USA  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Winpalm (Genentech)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/156,923

FILING DATE: 18-SEP-1998

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/359,705

FILING DATE: 20-DEC-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/286846

FILING DATE: 10-AUG-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/215139

FILING DATE: 18-MAR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: GENEENT.33CP2C1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 949/760-0404

TELEFAX: 949/760-9502

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1858 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear

US-09-156-923-7

Query Match 1.7%; Score 34.8; DB 3; Length 1858;  
Best Local Similarity 58.8%; Pred. No. 1.7;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 580 AACTGCTCTGTCAGAGGCGCTGCTTTGGTAGCTGTCTTCACCCAGTGGGTTTT 639  
DB 1215 AAGGGCTCTTGAGAGAACTGGCATGTATGCTGCTGCGTCGCCAGTGGGTTTT 1156

QY 640 GGAGAAGTCTCTTTGTTCCAGATGGCGAGTGGGCAT 681  
DB 1155 GGCAATGAGGTATAGTCCCATTTGTTGATGGGGGCTT 1114

RESULT 13  
US-08-359-705B-5/c

Sequence 5, Application US/08359705B  
Patent No. 584092

GENERAL INFORMATION:  
APPLICANT: Presta, Leonard G.

APPLICANT: Shelton, David L.

APPLICANT: Urfert, Roman

TITLE OF INVENTION: Human trk Receptors and Neurotrophic Factor Inhibitors

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California

COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Winpalm (Genentech)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/359,705B

FILING DATE: 20-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/286846

FILING DATE: 08/10/94

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/215139

FILING DATE: 03/18/94

ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD, Timothy E.

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P0873P2

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8674

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2715 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear

US-08-359-705B-5

Query Match 1.7%; Score 34.8; DB 2; Length 2715;  
Best Local Similarity 58.8%; Pred. No. 2.1;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 580 AACTGCTCTGTCAGAGGCGCTGCTTTGGTAGCTGTCTTCACCCAGTGGGTTTT 639  
DB 1215 AAGGGCTCTTGAGAGAACTGGCATGTATGCTGCTGCGTCGCCAGTGGGTTTT 1156  
QY 640 GGAGAAGTCTCTCTTTGTTCCAGATGGCGAGTGGGCAT 681  
DB 1155 GGCAATGAGGTATAGTCCCATTTGTTGATGGGGGCTT 1114

RESULT 14

US-08-286-846A-5/c  
Sequence 5, Application US/08286846A  
Patent No. 5877016

GENERAL INFORMATION:  
APPLICANT: Presta, Leonard G.

APPLICANT: Shelton, David L.



APPLICANT: Ufer, Roman  
TITLE OF INVENTION: Human trk Receptors and Neurotrophic Factor Inhibitors  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,846A  
FILING DATE: 05-Aug-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0873P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2715 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-286-846A-5

Query Match 1.7%; Score 34.8; DB 2; Length 2715;  
Best Local Similarity 58.8%; Pred. No. 2.1;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 580 AACTGCTGCTGCGCAGGGGCTGTTTGGTAGCCTTGTCCTCACCCACAGGCTTTT 639  
DB 1215 AAAGGGCTCTCTGAGAGAGTGGCCATTGATGGCTGGTGCCTCCAGTGGGTTT 1156  
QY 640 GGAGAGTCTCTGTTTCCAGATGGCGAGTGGGCGAT 681  
1155 GGCAATGAGGATATGATGCCATTGTTGATGGTGGGCTT 1114

## RESULT 15

US-08-457-880A-5/C  
Sequence 5, Application US/08457880A  
Patent No. 5910574  
GENERAL INFORMATION:  
APPLICANT: Leonard G. Presta  
APPLICANT: David L. Shelton  
APPLICANT: Roman Ufer  
TITLE OF INVENTION: HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,880A

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,622  
FILING DATE: 19-May-1995  
APPLICATION NUMBER: 08/286846  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Phd., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0873P1C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2715 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-457-880A-5

Query Match 1.7%; Score 34.8; DB 2; Length 2715;  
Best Local Similarity 58.8%; Pred. No. 2.1;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 580 AACTGCTGCTGCGCAGGGGCTGTTTGGTAGCCTTGTCCTCACCCACAGGCTTTT 639  
DB 1215 AAAGGGCTCTCTGAGAGAGTGGCCATTGATGGCTGGTGCCTCCAGTGGGTTT 1156  
QY 640 GGAGAGTCTCTGTTTCCAGATGGCGAGTGGGCGAT 681  
DB 1155 GGCAATGAGGATATGATGCCATTGTTGATGGTGGGCTT 1114

Search completed: October 21, 2002, 21:56:50  
Job time : 88 secs

**THIS PAGE BLANK (USPIC)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:11:18 : Search time 2477 Seconds  
(Without alignments)  
11459.062 Million cell updates/sec

Title: US-09-809-638-1  
Perfect score: 2103  
Sequence: 1 atgacctgcgtgtgagagaga.....ccaatacttttatgaac 2103

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Archived: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlinu:\*  
5: em\_estlov:\*  
6: em\_estlpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl1:\*  
10: gb\_estl2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vtc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736.2	35.0	1010	10	BG696417 602659516
2	524.8	25.0	872	10	BG247890 602359805
3	511	24.3	899	10	B1687517 603315072
4	492.4	23.4	748	10	B1854920 603381910
5	476.2	22.6	755	10	B1105048 602892888
6	456	21.7	585	10	B1343511 371695 MA
7	453.2	21.6	678	10	B1651562 603298034
8	445	21.2	596	10	B1343877 372176 MA
9	441.6	21.0	966	10	BG247983 602359913
10	436.6	20.8	597	9	AA710301 vt53b05.r
11	418	19.9	631	9	AA710301 vt53b05.x
12	411.8	19.5	528	9	AA710301 vt53b05.x
13	409.4	19.5	974	10	B1853011 603379834
14	395.8	18.8	617	9	A1962905 wt24f04.x
15	354.2	16.8	656	9	BB628092 BB628092
16	353.8	16.8	936	10	B1622283 601770979
17	346.4	16.5	534	9	AA572339 xq16f09.x

18	337.4	16.0	352	9	AA369925	AA369925 EST81558
19	332.4	15.8	378	10	B1001197	B1001197 PMO-HN007
20	330.6	15.7	672	9	B1181897	B1181897 B1181897
21	324	15.4	510	9	A1377551	A1377551 tc15h10.x
22	322.4	15.3	520	9	A1812045	A1812045 tw46e03.x
23	298.2	14.2	1079	10	B164016	B164016 601772165
24	286.4	13.6	485	9	A1373245	A1373245 q248b10.x
25	243	11.6	439	9	A1717137	A1717137 ut-r-y0-a
26	231.4	11.0	445	10	BE375574	BE375574 601225462
27	230.8	11.0	408	10	BE328934	BE328934 bq32d07.x
28	223	10.6	290	9	A1868470	A1868470 tc51d12.x
29	218.8	10.4	437	10	BG193159	BG193159 RST12284
30	202.2	9.6	394	9	A1826824	A1826824 wk56f12.x
31	185.4	8.8	313	10	BG971713	BG971713 EST-01 MO
32	180	8.6	495	9	AA863019	AA863019 0999f04.s
33	176.6	8.4	425	9	BB846675	BB846675 BB846675
34	175.4	8.3	428	9	BB845043	BB845043 BB845043
35	164.6	7.8	318	9	AA892501	AA892501 EST196304
36	140.4	6.7	370	9	AA523805	AA523805 n168h1.s
37	139.6	6.6	998	12	CNS078KT	AL434115 T3 end of
38	139.6	6.6	1115	12	CNS06PTX	AL408819 T7 end of
39	134.4	6.4	1056	12	CNS06MFW	AL405426 T3 end of
40	121.4	5.8	512	9	AA972717	AA972717 EST384809
41	117.8	5.6	1063	12	CNS06M3K	AL404982 T3 end of
42	112	5.3	305	9	AA195718	AA195718 zc33d07.s
43	110.6	5.2	174	10	BF543325	BF543325 ut-r-y0-a
44	108.6	5.2	482	9	BB744625	BB744625 BB744625
45	96.4	4.6	492	9	BB741693	BB741693 BB741693

#### ALIGNMENTS

RESULT 1  
LOCUS BG696417  
DEFINITION 602659516P1 NCI\_CGAP\_Skn3 Homo sapiens CDNA clone IMAGE:4802907 5',  
mRNA sequence.

ACCESSION BG696417  
VERSION BG696417.1  
KEYWORDS GI:13961539

SOURCE  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 1010)  
NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
COMMENT Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov

Tissue Procurement: James Cleaver, M.D.  
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LAM10697 row: O column: 04

High quality sequence stop: 803.  
location/Qualifiers

FEATURES  
SOURCE

1..1010  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="IMAGE:4802907"

/clone\_id="NCI\_CGAP\_Skn3"  
/lab\_host="DH10B (TL phage-resistant)"

/note="Organ: Skin; Vector: pcMV-SPOrt6; Site:1; Notti:  
Site:2; Salt: Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 318 a 208 c 230 g 254 t  
ORIGIN





## JOURNAL COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM12012 row: c column: 03  
 High quality sequence stop: 746.  
 Location/Qualifiers

## FEATURES

source

```

1..748
/organism="Mus musculus"
/strain="FVB/N"
/db.xref="taxon:10090"
/clone="IMAGE:5397050"
/clone_lib="NCL_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT    198 a    162 c    207 g    181 t
ORIGIN

```

Query Match 23.4%; Score 492.4; DB 10; Length 748;  
 Best Local Similarity 84.2%; Pred. No. 6.5e-127;  
 Matches 625; Conservative 0; Mismatches 111; Indels 6; Gaps 6;

```

OY 1189 CTGTGGCTCTTGTGGTGGATGTGTAGTACGACGATTAAGCCATGATGAG 1248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 CTGTGGCTCTTGTGGTGGATGTGTAGTACGACGATTAAGCCATGATGAG 60
OY 1249 AGAAAAAGTGGGCAAGTGGCACCACCAAGAGGTCTGTGCTGCTGCTTTCAG 1308
    || || || || || || || || || || || || || || || || || || || || ||
DB 61 AGCAGATTAGGCGGAGGCGCACAGGACAGATGCTGCGGCCATCTGCGCTTTAG 120
OY 1309 TTTGATATGACATGAGAGGCTGTCTAGTAAAGATCAGCTGCTCAATGAA 1368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 TTTGATATGACATGAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
OY 1369 ACAGGCGAGATTTATATACATTTTGGAGATGATCTTAAAGCCATATAGGGGAG 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 ACAGGCGAGATTTATATACCA-TTTGGAGAGCGATGCTTCAAGCCATATAGGGGAG 239
OY 1429 AATGACTTAACCATGTGCTAGGGGAAAAAGTTGGTTCTATACAGACTTTGTCACAG 1488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 240 AATGACTTAACCATGTGCTAGGGGAAAAAGTTGGTTCTATACAGACTTTGTCACAG 299
OY 1489 ACAAGGATACACTTGGGGGATTTATGCTTCAAGATACCAATTTGTGAATGTGAG 1548
    || || || || || || || || || || || || || || || || || || || || ||
DB 300 ACCAGG-ATCACACCTGGGGGATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 357
OY 1549 CATTCACCTTCCGTCACAGAGGGGAGATGCGACACGATCAGTACGACCTTAC 1608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 338 CATTCACCTTCCGTCACAGAGGGGAGATGCGACACGATCAGTACGACCTTAC 417
OY 1609 ATTTGGGCAAGCTGTGATTTTGTCTGACACACTTTGGGAGACACAGAGTACCTC 1668
    || || || || || || || || || || || || || || || || || || || || ||
DB 418 GTCTCAACAGAGCTGTGGA-TTTGTGTGTACACACTTTGGGAGATCATAGATGACTT 476
OY 1669 GACAGAGAACTCAGAGCTATTGCTGTTTAAACTACTGAAAAAGTACTCTATCAAGTG 1728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 477 GACAGAGAACTCAGAGCTATTGCTGTTTAAACTACTGAAAAAGTACTCTATCAAGTG 536
OY 1729 AATTTTGGGATATATCATCTTCAGAGCTGGCTCAGAGA-TTATCTACACTCAGCTA 1787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 537 AATTTTGGGATATATCATCTTCAGAGCTGGCTCAGAGATTTACATCACTCAGACAA 596

```

```

OY 1788 ACATGGCAATGTGAGAGATATGACAGA-GCAGTATCATGACAGATGCTGAATCATTA 1846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 597 ACATGGCAATGTGAGAGATATGACAGCGCTCAGATGAGAGAGATGCTGATCATCA 656
OY 1847 TGTATCGAGGCTGATCAGCTGGTGGTTATGCAAGATCTCCATGCTGATGATGAT 1906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 657 TGTACCGGGCTGTGATCAGCTGGGCTATGCAAGATCTCATGACAGATGATGACT 716
OY 1907 CAGAATTCAGATGCAAAATT 1928
    || || ||||| ||||| |||||
DB 717 CTGACATTCAGATGCGCAAAATT 738

```

RESULT 5  
 BI105048 755 bp mRNA linear EST 26-JUN-2001  
 LOCUS 602892888E1 NCL\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5037956 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI105048 GI:14555941  
 VERSION BI105048.1  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE  
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM1104 row: p column: 21  
 High quality sequence stop: 739.  
 Location/Qualifiers

## FEATURES

source

```

1..735
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="IMAGE:5037956"
/clone_lib="NCL_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary."
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT    176 a    165 c    205 g    209 t
ORIGIN

```

Query Match 22.6%; Score 476.2; DB 10; Length 755;  
 Best Local Similarity 81.1%; Pred. No. 2.3e-127;  
 Matches 614; Conservative 0; Mismatches 136; Indels 5; Gaps 5;

```

OY 683 CACATCCAGAGGCGACATCTTAACCATTTGAGAGTCACTGCTGCTGGCAAGTG 742
    || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 CCCACCCAGGCGCATCTATCTCTTTGGAGGAGGGCTTCTACGTGGCTCTCAAGTG 61
OY 743 GATTAGCTTCCATCTGTTTGTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTT 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 GACTAGTCTATCGGTTTCATCATGCTTCATGATGATGCTGTTAGCTGTGATGAGAG 121
OY 803 GAACAGCTTCACTCTCGGGGCTCTTACCTGACACAAAGGGCAGCTGCTGCTGCT 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 GAGCAGCTTGGCGCATGAGGTCTCTCTACCTGCGACAGAGGGCAGCTGCTGCTGAT 181

```







QY	914	TTATTAACTCAGGAGCAAAACCCCTGGGAAAACCATGACCATTCAGATATTTATCTTC	973
	1	TTGTGAACTCAGGAGAAAACCTCCGGAGAAGCAGATGCCCATGTATCCTTATATGTTT	60
QY	974	TAGAAATATTTTTCGTGCCCTGTGTGCACAGCTTTTAAGTTTGTCCAGAGAGTGTCTACG	1033
Db	61	TACAACACATTTTTCGTGCCCTGTGTGCACAGGAGTTTAA- TTGTTCGCCGAGAGTGTCTACG	119
QY	1034	CTAGAGAAGATACGATGTGCTTTTGGGGAGACAATGATTTAATATCGGCGTGAATATGC	1093
Db	120	CTAGAGAAAGATCTGATGTGCTTTTGGGGAGACAATATGTAATTAATTGATTTGATATGC	179
QY	1094	TATTGTTGCTTACAGAAAACCTTGACTTGTCTTTCAAACAAAACAGTTCTAAAGTGC	1153
Db	180	TGTTTGGAGACTTAAGAAAACCTTGACTTCTTCTTCAACAAAACAGTCTTAAAGGC	239
QY	1154	TTTTAGAAAAGGTAAATAATACATGAATCTTTTCTGTGGCTGCTTTGTGGTGGAT	1213
Db	240	TTTTAGATTCGACGTATTAATAATCATGAACTCATTTCTGTGGCTG- TTGTGTGTGGGCT	298
QY	1214	TGTTGGATTAGAGACTACGCGATAAAGCCTATAGAGAAAACCTGGGCAAAAGTGGCACCA	1273
Db	299	TACTGCGATTAGAGCCTACGCGACACAACTACAGAGGACAGTATAGCGAGGCGACAG	358
QY	1274	CCAAAGAGTCTCTGCTGCCCATCTGGCCCTTCAGTTTGGATATGACAAATGAAGCTGCT	1333
Db	359	CGAGCTGCTCTGCGCTGCCCATCTGGCCCTTTTAGG- TTGGATATGACAAATGAAGCTGCG	417
QY	1334	CTAGCTCTGAAAAGATACACACTACCTGCCTCAATGAAGACAGTGACAGATTCTAAACAATTT	1393
Db	418	CCAATCTAGAGAGGTCTGCTCTCAACGTCTCAAGGAGACAGGTGACATTTTATACACATTT	477
QY	1394	TGGAGAGTGATGCTTCTAAAGCCCTATATGGGGAGACAATGACTTAAACATGTGTGCTAGGG	1453
Db	478	TGGAGAGGAGGCTCTTCAAGCCCTATATAGGAGAACACGACTTAAACAATGTGTGCGGG	537
QY	1454	AAAAGTTGGGTTTCTATACAGACTTGTGGTCCAGACACAAAGGATATGACCTTGGGGGATTA	1513
Db	538	AGAACTGGGTTTCTACACAGACTTGTGGGCCAGACACAGG- ATACACACTGTGGGGGATTA	596
QY	1514	TGGCTTTGTCAAGATACCAATTTGTAATCTGAGACATCACCTTCTCCGTACACAGAG	1573
Db	597	TGGCTGCTGTGCGGTA- CGGATTTGTCAATGTGAGGACATCACCTTCTCCGTGCGCAGAGN	655
QY	1574	G 1574	
Db	656	G 656	
RESULT 10			
AA710301			
LOCUS	AA710301	597 bp	mRNA linear EST 24-DEC-1997
DEFINITION	v13b05.1 Barstead mouse irradiated colon MPlRB7 Mus musculus cDNA		
KEYWORDS	clone IMAGE:1166769 5' similar to TR:P/8781 P/8781 F1SSION YEAST ; ,		
ACCESSION	AA710301		
VERSION	AA710301.1 GI:2720219		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		

## COMMENT

Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:632681

Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 509.

## FEATURES

source

Location/Qualifiers

1..597

/organism="Mus musculus"

/strain="FVB/N"

/db.xref="taxon:10090"

/clone="IMAGE:1166769"

/clone\_lib="Barstead mouse irradiated colon MCLRb7"

/dev\_stage="8 weeks"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (pharmacia) with a modified  
 polylinker. Site\_1: EcoRI. Site\_2: NotI. Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 15'GCTTACGATCGAGTGGAGCGCCGCCCTTTTCTTTTCTTTTCTTTT  
 T 3'; double-stranded cDNA was ligated to Eco RI  
 adaptors (AATTCGATCCTTG), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead."

BASE COUNT 174 a 141 c 144 g 138 t  
 ORIGIN

Query Match 20.8%; Score 436.6; DB 9; Length 597;  
 Best Local Similarity 84.6%; Pred. No. 2.7e-111;

Matches 490; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

## FEATURES

source

1..631

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="IMAGE:2694176"

/clone\_lib="NCI-CGAP\_K1d11"

/lab\_host="DH10B"

/note="Organ: Kidney. Vector: pT73D-Pac (pharmacia) with  
 a modified polylinker. Site\_1: Not I; Site\_2: EcoRI;  
 Plasmid DNA from the normalized library NCI-CGAP Kid3 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (clonoids 1323376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 168 a 116 c 101 g 245 t 1 others  
 ORIGIN

Query Match 19.9%; Score 418; DB 9; Length 631;  
 Best Local Similarity 97.5%; Pred. No. 4.7e-106;

Matches 424; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2009 ATTTAATCCAGATTGGATCTCAAGAAGACACA 2047  
 Db 559 ACTTCAACCCAGATTGGCTCTCAAGAAGACACA 597

## RESULT 11

AM235545/C

LOCUS

DEFINITION

AM235545.1 x1 CGAP Kid11 Homo sapiens cDNA clone IMAGE:2694176 3'  
 similar to SW:Y02\_YEAST P25618 HYPOTHETICAL 107.9 KD PROTEIN IN  
 POL4-SRD1 INTERGENIC REGION. ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: 40UP from Gibco  
 High quality sequence stop: 441.

## FEATURES

source

1..631

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="IMAGE:2694176"

/clone\_lib="NCI-CGAP\_K1d11"

/lab\_host="DH10B"

/note="Organ: Kidney. Vector: pT73D-Pac (pharmacia) with  
 a modified polylinker. Site\_1: Not I; Site\_2: EcoRI;  
 Plasmid DNA from the normalized library NCI-CGAP Kid3 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (clonoids 1323376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 168 a 116 c 101 g 245 t 1 others  
 ORIGIN

Query Match 19.9%; Score 418; DB 9; Length 631;  
 Best Local Similarity 97.5%; Pred. No. 4.7e-106;

Matches 424; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 1669 GACAGAAACCTGACGCTATTCCTTTCAAACTAGCAAGAAAGTACTATCACTG 1728

Db 631 GACAGAAAAATCAGACGCTATTCCTTTCAAACTAGCAAGAAAGTACTATCACTG 572

Oy 1729 ATATTTCGGGATATATCACTTCACACCTGGCTCCAGAGATTATCTACGCTACTGA 1788

Db 571 ATATTTCGGGATATATCACTTCACACCTGGCTCCAGAGATTATCTACGCTACTGA 512

Oy 1789 CATGGCAATGTAAAGATATGACAGCACTGATCATGACAGATGGTGAATATATATG 1848

Db 511 CATGGCAATGTAAAGATATGACAGCACTGATCATGACAGATGGTGAATATATATG 452

Oy 1849 TATCAGAGGCTGATCAGGTGGGTTATGCAAGAAATCTCCATGTAAGTGAATGTA 1908

Db 499 CCAATTACAGAGACAAGAAAGTGTCTATGACACCGAGAGAGTTCACCAAGAACTTC 558

```

|||||
Db 451 TATCAGGGCTGATCAGGTGGGTATGCAAGATCTCCATGCTGAGACTGATGATTC 392
Oy 1909 GAAATTCAGATGCAAAATTTTATAGTATCCCTGATGACCCACATTAATATAGACACACG 1968
Db 391 GAAATTCAGATGCAAAATTTTATAGTATCCCTGATGACCCACATTAATATAGACACACG 332
Oy 1969 AAGTGTGATGACACAGAGAAATTTCTGAGAAATTTATTTATCCAGATTTGGA 2028
Db 331 AAGTGTGATGACACAGAGAAATTTCTGAGAAATTTATTTATCCAGATTTGGA 272
Oy 2029 TCTTACAAAGAGACACATTTATGAAAACACATCTTTTCATATGATATCTCCAA 2088
Db 271 TCTTACAAAGAGACACATTTATGAAAACACATCTTTTCATATGATATCTCCAA 212
Oy 2089 TACTTTTATGAAAC 2103
211 TACTTTTATGAAAC 197

RESULT 12
LOCUS BE180939 528 bp mRNA linear EST 22-JUN-2000
DEFINITION PM3-HT0627-270300-001-e03 HT0627 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE180939
VERSION BE180939.1 GI:8660115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 528)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Goi,dman,G.H., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ruda Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-PM3-HT0627-270
300-001-e03&t3=2000-03-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 2
High quality sequence stop: 475.
FEATURES
Source
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0627"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site:1: Sma1;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 151 a 114 c 131 g 132 t
ORIGIN

```

```

Query Match 19.6%; Score 411.8; DB 9; Length 528;
Best Local Similarity 95.1%; Pred. No. 2,4e-104;
Matches 490; Conservative 0; Mismatches 17; Indels 8; Gaps 6;

Oy 1218 GGGATTAGGACTAGCGCATTAAGCCCTATAGAGAGAAAGTGGCAAGTGGACACCA 1277
Db 3 GGGATTAGGACTAGCGCATTAAGCCCTATAGAGAGAAAGTGGCAAGTGGACACCA 59
Oy 1278 AGAGGTCTCTGCTGACATCTGGCCCTTCAGTTTGATATGACATGAGAGGTGGTCTAG 1337
Db 60 AGAGGTCTCTGCTGACATCTGGCCCTTCAGTTTGATATGACATGAGAGGTGGTCTAG 119
Oy 1338 TCTAGAAAGATCAGCTCAGCTCCTCAATGAAGAGTGCAGATTTCAATGAATTTTGA 1397
Db 120 TCTAGAAAGATCAGCTCAGCTCCTCAATGAAGAGTGCAGATTTCAATGAATTTTGA 179
Oy 1398 GAGTGATGCTTCTAAGCCCTATATGGGAGACAAATGACTTAAACATGTGGTAGGAGAAA 1457
Db 180 GAGTGATGCTTCTAAGCCCTATATGGGAGACAAATGACTTAAACATGTGGTAGGAGAAA 239
Oy 1458 GTTGGGTTTCTATACAGACTTTGGTCCAGACACAGATATCACTTTGGGGATTATGCG 1517
Db 240 GTTGGGTTTCTATACAGACTTTGGTCCAGACACAGATATCACTTTGGGGATTATGCG 299
Oy 1518 TTTGTCAAGATACCCCAATTTGTAATCTGAGCATCACCTTCTCCGTACCAAGAGGCGGA 1577
Db 300 TTTGTCAAGATACCCCAATTTGTAATCTGAGCATCACCTTCTCCGTACCAAGAGGCGGA 359
Oy 1578 GATCGCACAGCCATACATTTGA--CCGTTAACATTTGGGCGCA--GCTGGTGGATTTTGT 1634
Db 360 GATCGCACAGCCATACATTTGA--CCGTTAACATTTGGGCGCA--GCTGGTGGATTTTGT 419
Oy 1635 CGTGACACACTTTGGGAACCA--CGAAGATACCTGACAGAGAAATGCG--AGGCTTTTCT 1692
Db 420 CGTGACACACTTTGGGAACCA--CGAAGATACCTGACAGAGAAATGCG--AGGCTTTTCT 479
Oy 1693 GTTTCAAAGTACTGAAAGTACTCTATATCAAGT 1727
Db 480 GTTTCAAAGTACTGAAAGTACTCTATATCAAGT 514

RESULT 13
LOCUS B1853011 974 bp mRNA linear EST 10-OCT-2001
DEFINITION 603379834F1 NIH_CGAP_Mam3 Mus musculus CDNA clone IMAGE:5389784 5',
ACCESSION B1853011
VERSION B1853011.1 GI:15993758
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE 1 (bases 1 to 974)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov
Tissue Procurement: Lothar Henninghausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LMNL1993 row: d column: 09
High quality sequence stop: 765.
FEATURES
Source
Location/Qualifiers
1..974
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5389784"

```

/clone\_lib="NIH\_CGAP\_Mam3"  
 /tissue\_type="tumor, gross tissue"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert 2 kb. Library constructed by Life  
 Technologies, catalog #12017-018. Investigators providing  
 samples: Lothar Hennigshausen/Chu-Xia Deng, NIH Reference  
 for transgenic model: Xu et al., Nature Genetics 22, 37-43  
 (1999). Note: this is a NCI\_CGAP Library."

BASE COUNT 198 a 272 c 274 g 230 t  
 ORIGIN

Query Match 19.5%; Score 409.4; DB 10; Length 974;  
 Best Local Similarity 78.1%; Pred. No. 1.5e-103;  
 Matches 505; Conservative 0; Mismatches 141; Indels 1; Gaps 1;

OY 1 ATGACCTCGCTGTGGAGAGAAATCTTGTGAGTGCCTGGTGGAGATGTTTCTTGCTCT 60  
 DB 123 ATGCCAGCGCTGTGGAGAGCAATTCCTGGAGACCTTCTGGCTACGCTCTGCTCT 182  
 OY 61 CTTCTACATGACCTGGAGCCGATGATCTATCTTCTTCTTCCAAACACTAGAACTCACT 120  
 DB 183 CTTCTACAGGAGCTAGAGCCCATGATCTATTTATTTCCCTCCAAACGCTGAGCTCACA 242  
 OY 121 GGGCTTGAAGTTTATGATGATCTTCTTCCAAATTCCTCAACAATTAATCTCTTTC 180  
 DB 243 GGGCTGGAATTTCTTTGTGTACCTTTCTCTCCGATACCTTTAAACAATTTCCCCACTC 302-  
 OY 181 TGGAAATGTTTACACAGAGTGTATGCTTACCCCTGCTGAGAGTAATCACTATTTGGCAGC 240  
 DB 303 TGGAAATGCTCAATTAAGAGTGAACACTCTCTCTGCTGAGAGTATGCTACTAGGCAAC 362  
 OY 241 ATAGCTCTCTCCAGCTCCCAATGCCAACTGCATGATGATGCTTCTGCTTGGGCTG 300  
 DB 363 ATAGCTCTCTTGGAGCTCCCAATGCTTAAGCTTCACTGATGATGCTCTGCTGCTGCTG 422  
 OY 301 TCTTCTCTCACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 DB 423 TCTTCTCTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 482  
 OY 361 CTCAGATTTGGGGATTCATTTTAGAGACAGATTTTCTTGTCTTCTTCTTCTTCTTCT 420  
 DB 483 CTCAGATTTGGGGATTCATTTTAGAGACAGATTTTCTTGTCTTCTTCTTCTTCTTCT 542  
 OY 421 ACTTCACTAAACCAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 543 ACATCACTAAACCAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 602  
 OY 481 GCCATAGCCACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 DB 603 GCCATAGCCACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662  
 OY 541 AAGACTGTTGAGGTAGCCAGGGG-ATGGCTCTTAGACCCCAATGCTGCTGCGAGGGGC 599  
 DB 663 AAGCCCGGTAGGTAGCCAGGGGAGGAGCCCTCTATCTAGTGTGGCTTCCAGGCGC 722  
 OY 600 TGCCTTGGTACCTTGTGTTCTCCACCCACTGGGTTTGGAGAG 646  
 DB 723 CGCTTTGGAGCCCTCTGTTTCTTAACCTGATGATCTTGGAGAG 769

RESULT 14  
 A1962905/c 617 bp mRNA linear EST 20-AUG-1999  
 LOCUS A1962905  
 DEFINITION w124f04.x1 NCI\_CGAP\_Ut1 Homo sapiens cDNA clone IMAGE:2508415 3'  
 similar to TR:P78781 P78781 FLSSION YEAST ;, mRNA sequence.  
 ACCESSION A1962905  
 VERSION A1962905  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 617)  
 AUTHORS NCI\_CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI\_CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: -400P from GIDCO  
 High quality sequence stop: 407.  
 Location/Qualifiers

## FEATURES

source 1. .617  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2508415"  
 /clone\_lib="NCI\_CGAP\_Ut1"  
 /tissue\_type="well-differentiated endometrial  
 adenocarcinoma, 7 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Life Technologies catalog #:  
 11538-014"  
 BASE COUNT 159 a 109 c 102 g 246 t 1 others  
 ORIGIN

Query Match 18.8%; Score 395.8; DB 9; Length 617;  
 Best Local Similarity 99.0%; Pred. No. 7.9e-100;  
 Matches 408; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 1693 GTTTCAAACACTGTAAGTAGCTCTAATCAAGATATTTCGGATATATCACTTCA 1752  
 DB 617 GTTTCAAACACTGTAAGTAGCTCTAATCAAGATATTTCGGATATATCACTTCA 558  
 OY 1753 GCACCTGGCTCCAGAGATATCTACAGCTCACTGAATGCAATGGAAGATATGAG 1812  
 DB 557 GCACCTGGCTCCAGAGATATCTACAGCTCACTGAATGCAATGGAAGATATGAG 498  
 OY 1813 AGCATGATCATGACAGATGCTGTAATCATATGATATGATGATGATGATGATGAT 1872  
 DB 497 AGCATGATCATGACAGATGCTGTAATCATATGATATGATGATGATGATGATGAT 438  
 OY 1873 TATGCAAGAATCTCCCATGCTGTAATGATGATGATGATGATGATGATGATGATGAT 1932  
 DB 437 TATGCAAGAATCTCCCATGCTGTAATGATGATGATGATGATGATGATGATGATGAT 378  
 OY 1933 ATCCCTGATGACCCCATTAATATAGAGACAACGAAAGTGTATAGACACAGAGAA 1992  
 DB 377 ATCCCTGATGACCCCATTAATATAGAGACAACGAAAGTGTATAGACACAGAGAA 318  
 OY 1993 GTTTCGAGAAATTCATTTTATTCCTCCAGATTTGGATCTCTCAAAAGAGACACATTA 2052  
 DB 317 GTTTCGAGAAATTCATTTTATTCCTCCAGATTTGGATCTCTCAAAAGAGACACATTA 258  
 OY 2053 GAAACACACATATTTTCATATGAAATCT-CCCAATATCTTTTATGAAC 2103  
 DB 257 GAAACACACATATTTTCATATGAAATCTCCCAATATCTTTTATGAAC 206

RESULT 15  
 BB628092 656 bp mRNA linear EST 31-AUG-2001  
 LOCUS BB628092  
 DEFINITION BB628092 RIKEN full-length enriched, adult male urinary bladder Mus  
 musculus cDNA clone 9530064017 5', mRNA sequence.  
 ACCESSION BB628092



**THIS PAGE BLANK (USPTO)**

## RESULT 1

Q9H720

ID Q9H720

PRELIMINARY;

PRT: 699 AA.

AC Q9H720;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE CDNA: FLJ21511 fis, clone COL05748.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=COLON;

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK025164; BAB15080.1; -.

SQ SEQUENCE 699 AA; 78565 MW; 2F839563189A0523 CRC64;

Query Match 99.8%; Score 3713; DB 4; Length 699;

Best Local Similarity 99.9%; Pred. No. 3.6e-263;

Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSWLREILLESLLGCVSWSLYHDLGPMIYYFPLQTLTLELGFSIAFLSPIFLTITPF 60  
 Db 1 MTSWLREILLESLLGCVSWSLYHDLGPMIYYFPLQTLTLELGFSIAFLSPIFLTITPF 60

Qy 61 WKLNVKKWMLTLRLIITIGSIASFQAPNAKRLMLVLCVSSSLIVQAVTWWSGSHLQRY 120  
 Db 61 WKLNVKKWMLTLRLIITIGSIASFQAPNAKRLMLVLCVSSSLIVQAVTWWSGSHLQRY 120

Qy 121 LRIWGFILGQIVLVVLRWIYTSLNPIWSYQMSNKVILTLSAATLDRICTDGDCKSPEEK 180  
 Db 121 LRIWGFILGQIVLVVLRWIYTSLNPIWSYQMSNKVILTLSAATLDRICTDGDCKSPEEK 180

Qy 181 KTGEVATGMASRPNWLLAGAAGFSLVFLTHWVFEVSLVSRWAVSGHPHGPDPNPF 240  
 Db 181 KTGEVATGMASRPNWLLAGAAGFSLVFLTHWVFEVSLVSRWAVSGHPHGPDPNPF 240

Qy 241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTAS 300  
 Db 241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTAS 300

Qy 301 TLGHLINSGTNPCKMTIAMIFYLLEIFFCWCTAFKFVPGGVYARERSDVLLGTMM 360  
 Db 301 TLGHLINSGTNPCKMTIAMIFYLLEIFFCWCTAFKFVPGGVYARERSDVLLGTMM 360

Qy 361 GLNMLFGPKKNLDLLQTKNSSKVLFRKSEKYMKFLWLLVGVGLLGLRHKAYERKLG 420  
 Db 361 GLNMLFGPKKNLDLLQTKNSSKVLFRKSEKYMKFLWLLVGVGLLGLRHKAYERKLG 420

Qy 421 KVAPTKEVSAAIWPFRTGYDNEGWSLERSAHLNETGADFITILESASKPYMGNDLT 480  
 Db 421 KVAPTKEVSAAIWPFRTGYDNEGWSLERSAHLNETGADFITILESASKPYMGNDLT 480

Qy 481 MWLGEKLGFTDFGPSTRYHTWGIMALSRYPIVKSEHLLPSPEGEIAPAITLTVNIS 540  
 Db 481 MWLGEKLGFTDFGPSTRYHTWGIMALSRYPIVKSEHLLPSPEGEIAPAITLTVNIS 540

Qy 541 LVDFVVTHTFGNHEDDLDRKLQAIKSKLLSSNQVIFLGYITSAPGSRDYQLTEHGN 600  
 Db 541 LVDFVVTHTFGNHEDDLDRKLQAIKSKLLSSNQVIFLGYITSAPGSRDYQLTEHGN 600

Qy 601 KDIDSTDHWRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQ 660  
 Db 601 KDIDSTDHWRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQ 660

Qy 661 DHREVSEKIHFNPRFGSYKEGHNYENNHNHFMNTPKYFL 699  
 Db 661 DHREVSEKIHFNPRFGSYKEGHNYENNHNHFMNTPKYFL 699

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 21:56:53 ; Search time 63 Seconds  
(without alignments)  
1919.420 Million cell updates/sec

Title: US-09-809-638-2

Perfect score: 3720

Sequence: 1 MTSLWREILLESLLGCVSWS.....EGHNYENNHHFHMTPKVEL 699

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriaph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3713	99.8	699	4 Q9H720	Q9H720 homo sapien
2	3125	84.0	699	11 Q91YL7	Q91YL7 mus musculus
3	639	17.2	425	3 Q9P6M4	Q9P6M4 schizosacch
4	625.5	16.8	401	3 P78781	P78781 schizosacch
5	256	6.9	688	3 Q9HD22	Q9HD22 schizosacch
6	141.5	3.8	279	2 Q9S1Z5	Q9S1Z5 streptomyce
7	132	3.5	583	8 Q950T2	Q950T2 hyaloraphid
8	128.5	3.5	396	2 P95556	P95556 pseudomonas
9	124	3.3	609	8 Q79569	Q79569 mustelus ma
10	123.5	3.3	269	11 Q91VY8	Q91VY8 mus musculus
11	122	3.3	636	6 Q9GJX6	Q9GJX6 sus scrofa
12	122	3.3	643	6 Q9GJX7	Q9GJX7 sus scrofa
13	122	3.3	766	17 Q9DYP5	Q9DYP5 pyrococcus
14	121.5	3.3	591	8 Q36149	Q36149 trachemys s
15	121	3.3	605	8 Q9B6Y4	Q9B6Y4 casuarinus b
16	119	3.2	427	16 O53515	O53515 mycobacteri

17	119	3.2	612	8 Q9G6Q3	Q9G6Q3 diplophos t
18	118.5	3.2	449	2 Q9RL05	Q9RL05 streptomyce
19	118.5	3.2	605	8 Q957Z0	Q957Z0 casuarinus c
20	118.5	3.2	613	8 Q94SY9	Q94SY9 crenimugil
21	118.5	3.2	694	2 Q9L179	Q9L179 streptomyce
22	117.5	3.2	583	5 Q17069	Q17069 caenorhabdi
23	117	3.1	604	8 Q9B6Z4	Q9B6Z4 apteryx man
24	116.5	3.1	570	8 Q9B510	Q9B510 tetradontop
25	116.5	3.1	613	8 Q94S29	Q94S29 mugil cepha
26	116	3.1	605	8 Q9B6W0	Q9B6W0 dinornis gl
27	116	3.1	605	8 Q9B6U8	Q9B6U8 dromaius no
28	116	3.1	605	8 Q958A0	Q958A0 dromaius no
29	115.5	3.1	590	17 Q96XC0	Q96XC0 sulfolobus
30	115	3.1	317	16 Q97K74	Q97K74 cicostridium
31	115	3.1	603	8 Q9MR50	Q9MR50 ciconia cic
32	115	3.1	603	8 Q958D4	Q958D4 tinamus maj
33	114.5	3.1	603	8 Q9GA20	Q9GA20 tupaia beia
34	114.5	3.1	829	10 Q9LMD1	Q9LMD1 arabidopsis
35	114	3.1	395	2 Q9Z404	Q9Z404 pseudomonas
36	114	3.1	606	8 Q9TA19	Q9TA19 loxodonta a
37	114	3.1	1201	16 Q9HU70	Q9HU70 pseudomonas
38	113.5	3.1	276	2 Q9ZBN3	Q9ZBN3 streptomyce
39	113.5	3.1	605	8 Q957X8	Q957X8 anomalopter
40	113	3.0	604	8 Q9B6X2	Q9B6X2 crypturellu
41	113	3.0	689	13 Q9W701	Q9W701 xenopus lae
42	112	3.0	438	10 Q9ZU86	Q9ZU86 arabidopsis
43	111.5	3.0	562	16 Q9ZT98	Q9ZT98 rhizobium m
44	111	3.0	544	17 Q9HQN5	Q9HQN5 halobacteri
45	111	3.0	614	8 Q94YP6	Q94YP6 osteoglossu

## ALIGNMENTS

### RESULT 1

Q9H720	PRELIMINARY;	PRT;	699 AA.
ID	Q9H720		
AC	Q9H720;		
DC	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DE	CDNA: FLJ21511 FIS, CLONE COL05748.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=COLON;		
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,		
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK025164; BAB15080.1; -		
SQ	SEQUENCE 699 AA; 78565 MW; 2F839563189A0523 CRC64;		

Query Match 99.8%; Score 3713; DB 4; Length 699;  
Best Local Similarity 99.9%; Pred. No. 2.7e-266;  
Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTSLWREILLESLLGCVSWSLYHDLGPMIYFFPLOTLELTGLGEGFSTAFSLPFLITTPF	60
Db	1	MTSLWREILLESLLGCVSWSLYHDLGPMIYFFPLOTLELTGLGEGFSTAFSLPFLITTPF	60
QY	61	WKLVNKKWMLTLRIITIGSIASFQAPNAKRLMLVLAGVSSSLIVQAVTWSSGSHLQRY	120
Db	61	WKLVNKKWMLTLRIITIGSIASFQAPNAKRLMLVLAGVSSSLIVQAVTWSSGSHLQRY	120
QY	121	LRWGFILGQIVLVLRWYTSNLPWYSQMSKNKVIITLSAIALTLDRIGTDCGSKPEK	180
Db	121	LRWGFILGQIVLVLRWYTSNLPWYSQMSKNKVIITLSAIALTLDRIGTDCGSKPEK	180

QY 181 KTGEVATGMSRPNWLLAGAAFGSLVFLTHWVGEVSLVSRMAVSGHPHPDPNPFGGA 240  
 Db 181 KTGEVATGMSRPNWLLAGAAFGSLVFLTHWVGEVSLVSRMAVSGHPHPDPNPFGGA 240  
 QY 241 VLLCLASGLMLPSCLEWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAITFASWMPQ 300  
 Db 241 VLLCLASGLMLPSCLEWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAITFASWMPQ 300  
 QY 301 TLGHLINSNTNPKTMTIAMIFYLLEFFCAWCTAFKFPVGGVYARERSDVLGTMMLII 360  
 Db 301 TLGHLINSNTNPKTMTIAMIFYLLEFFCAWCTAFKFPVGGVYARERSDVLGTMMLII 360  
 QY 361 GLNMLFGPKKNDLLQTKNSKLVPRKSEKYMFLWLLVGVGLGLGLRHKAYERKLG 420  
 Db 361 GLNMLFGPKKNDLLQTKNSKLVPRKSEKYMFLWLLVGVGLGLGLRHKAYERKLG 420  
 QY 421 KVAPTKEVSAAIWPPFRGYDNEGWSLERSAHLNETGADFITILES DASKPYMGNNDLT 480  
 Db 421 KVAPTKEVSAAIWPPFRGYDNEGWSLERSAHLNETGADFITILES DASKPYMGNNDLT 480  
 QY 481 MWLGEKLGFTYDFGSPSTRYHTWGMALSRYPYIVKSEHLLPSPEGETAPAITLTVNISGK 540  
 Db 481 MWLGEKLGFTYDFGSPSTRYHTWGMALSRYPYIVKSEHLLPSPEGETAPAITLTVNISGK 540  
 QY 541 LVDVVTTHFGNHEDDLDRKLAIAVSKLLKSSNQVIFLGYITSAPGSRDYQLQTEHGNV 600  
 Db 541 LVDVVTTHFGNHEDDLDRKLAIAVSKLLKSSNQVIFLGYITSAPGSRDYQLQTEHGNV 600  
 QY 601 KDIDSTDHRCWEYIMYRGLIRGLYARISHAELSDSEIQMAKFRIPDDPTNYRDNOKVVI 660  
 Db 601 KDIDSTDHRCWEYIMYRGLIRGLYARISHAELSDSEIQMAKFRIPDDPTNYRDNOKVVI 660  
 QY 661 DHRVSEKIHFNPRFGSYKEGHNYENNHHFMMNTPKYFL 699  
 Db 661 DHRVSEKIHFNPRFGSYKEGHNYENNHHFMMNTPKYFL 699

## RESULT 2

Q91YL7 PRELIMINARY; PRT; 699 AA.  
 AC Q91YL7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL TO HYPOTHETICAL PROTEIN FLJ21511.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC016523; AAH16523.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 699 AA; 78179 MW; 4F41975D6570D5F8 CRC64;

Query Match 84.0%; Score 3125; DB 11; Length 699;  
 Best Local Similarity 82.8%; Pred. No. 8.5e-223;  
 Matches 579; Conservative 41; Mismatches 79; Indels 0; Gaps 0;

QY 1 MTSILWREILLESILGCVSWSLYHDLGPMIYFFLPQTLTLETLGEGFSIAFLSPILITTF 60  
 Db 1 MGLWRAIAETLLGVYVSWSLYHGLSPMIYFFLPQTLTLETLGEGFCVAFILSPILITPL 60  
 QY 61 WKLNVKMMTLRITIGSIASQAPNAKRLRMVLVIALGVSSSLIVQAVTWWSGSLQRY 120  
 Db 61 WKLNVKMMTLRITIGSIASQAPNAKRLRMVLVIALGVSSSLIVQAVTWWSGSLQRY 120  
 QY 121 LRTWGTILGOVLVLRITWTSNPIWSYOMSKVILTLSAIALDRICTGDCSPKEK 180  
 Db 121 LRTWGTILGOVLVLRITWTSNPIWSYOMSKVILTLSAIALDRICTGDCSPKEK 180

QY 181 KTGEVATGMSRPNWLLAGAAFGSLVFLTHWVGEVSLVSRMAVSGHPHPDPNPFGGA 240  
 Db 181 KPREVATGRTSLSSWLLPGAAGFSLFLTHWIFGEVSIIVSRMAVSGHPHPDPNPFGGA 240  
 QY 241 VLLCLASGLMLPSCLEWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAITFASWMPQ 300  
 Db 241 VLLCLASGLMLPSCLEWFRGTGLIWWVTGTASAAGLLYLHTWAAVSGCVFAITFASWMPQ 300  
 QY 301 TLGHLINSNTNPKTMTIAMIFYLLEFFCAWCTAFKFPVGGVYARERSDVLGTMMLII 360  
 Db 301 TLGHLINSNTNPKTMTIAMIFYLLEFFCAWCTAFKFPVGGVYARERSDVLGTMMLII 360  
 QY 361 GLNMLFGPKKNDLLQTKNSKLVPRKSEKYMFLWLLVGVGLGLGLRHKAYERKLG 420  
 Db 361 GLSMLFGPKKNDLLQTKNSKLVPRKSEKYMFLWLLVGVGLGLGLRHKAYERKLG 420  
 QY 421 KVAPTKEVSAAIWPPFRGYDNEGWSLERSAHLNETGADFITILES DASKPYMGNNDLT 480  
 Db 421 RGAPATVVSAAIWPPFRGYDNEGWSLERSAHLNETGADFITILES DASKPYMGNNDLT 480  
 QY 481 MWLGEKLGFTYDFGSPSTRYHTWGMALSRYPYIVKSEHLLPSPEGETAPAITLTVNISGK 540  
 Db 481 MWLGEKLGFTYDFGSPSTRYHTWGMALSRYPYIVKSEHLLPSPEGETAPAITLTVNISGK 540  
 QY 541 LVDVVTTHFGNHEDDLDRKLAIAVSKLLKSSNQVIFLGYITSAPGSRDYQLQTEHGNV 600  
 Db 541 LVDVVTTHFGNHEDDLDRKLAIAVSKLLKSSNQVIFLGYITSAPGSRDYQLQTEHGNV 600  
 QY 601 KDIDSTDHRCWEYIMYRGLIRGLYARISHAELSDSEIQMAKFRIPDDPTNYRDNOKVVI 660  
 Db 601 KDIDSTDHRCWEYIMYRGLIRGLYARISHAELSDSEIQMAKFRIPDDPTNYRDNOKVVI 660  
 QY 661 DHRVSEKIHFNPRFGSYKEGHNYENNHHFMMNTPKYFL 699  
 Db 661 DHRVSEKIHFNPRFGSYKEGHNYENNHHFMMNTPKYFL 699

## RESULT 3

Q9P6M4 PRELIMINARY; PRT; 425 AA.  
 AC Q9P6M4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 48.6 KDA PROTEIN (FRAGMENT).  
 GN SPAC688.01.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Aert R., Robben J., Volckaert G., Wood V., Rajandream M.A.,  
 RA Barrell B.G.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL355632; CAB90768.1;  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 425 AA; 48575 MW; 1B17FE034278A0E4 CRC64;

Query Match 17.2%; Score 639; DB 3; Length 425;  
 Best Local Similarity 34.2%; Pred. No. 3.6e-39;  
 Matches 152; Conservative 70; Mismatches 178; Indels 44; Gaps 10;

QY 276 LYL-HTWAAVSGCVFAITF-----ASMMPQTLGHLINSNTNPKTMTIAMIFYLLEIFF 329  
 Db 4 MYINHCWCSVGLGLIFTSVLYIYFSASI-----RISSFYSPAKVWGGAFVLYVLSLA 56  
 QY 330 CAWCTAFKFPVGGVYARERSDVLGTMMLIIGLNMFLF-----GPKNLDLLQTK 379  
 Db 57 HWVVAIEFVPGPILRRTSYIL----IFIGNLAALVPAYSGESKEPKADSSVYDIK 112

us-09-809-638-2.rspt

Dd	240	HATIDVYGEIIDLIVVYSHNGQYESQLDRRLQSTELARIMRESPTLVGLGVVSWVGEPQ	299
Oy	552	LQLEHGNVKRIDSTDHDRCWEETMYKLLILGTAARISHAELSSEIQMAKFRIIPDDPTN	655
Dd	300	TTLTRDTMDLIEPDAVDYRMOQITFYRRVKRIGGARLRHSRTITLOTGTFLVTKDL-	357
Oy	652	YRDNOKVVIDHREVSEKIHFMPNF-GSYKESEHNENNHHFMMNPXYE	698
Dd	358	---GNVRIDKENHPESHRYPSLEGTEGVNHGYTDNNLVH--EPWYY	400
 RESULT 5 O9HDZ2 ID O9HDZ2 PRELIMINARY; PRT; 688 AA. AC O9HDZ2: Dd - 01-Mar-2001 (TREMBLrel_16, Created) Dt 01-Mar-2001 (TREMBLrel_16, Last sequence update) Dt 01-Mar-2001 (TREMBLrel_16, Last annotation update) De HYPOTHETICAL 77.5 KDA PROTEIN (FRAGMENT). GN SPAC589.12. OS Schizosaccharomyces pombe (fission yeast). OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; CC Schizosaccharomycetales; Schizosaccharomycetaceae; OC Schizosaccharomycetes.			

RP	SEQUENCE FROM N.A.
RC	STRAIN-972H:-
RA	Wood V., Rajadream M.A., Barrell B.G., Aert R., Robben J.,
RA	Welters I., Grymoprez B., Volckert G.,
RL	Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AL512496; CAC19769.1; -
KM	Hypothetical protein.
FT	NON TER
SO	SEQUENCE 688 AA; 77474 MW; 3985EAC23CDA34FF CRC64;
Query Match	6.9%; Score 256; DB 3; Length 688;
Best Local Similarity	22.5%; Pred. No. 1.5e-10;
Matches 97; Conservative 77; Mismatches 197; Indels 60; Gaps 11	
QY	11 ESLLGCVSMLKHDIGPMIVPEPLDTLELTGLEGFSAIFSLPIFLITPFPVKLVNKKMWL 70
DB	285 EYLSLFVPSVILTSGLLWVFPLMHMGISGYEACTIFELSPDLGIPLLRFKASVVPVI 344
QY	71 TLLRIITIGSIASFQAPNAKLRLMYLALGVSSSLIYQAATWW-----SGSHLQRYLRIM 124
DB	345 FL-FLNVIGIAAYKLIEDPVHRLFVTAF----SVCECLAMWSLFSFNIPENAIERKS 398
QY	125 GTLDGIYLVYRIRMTSLNPWT-----SYQMSNKYLILISAATLIDRTGTDDGCK 176
DB	399 TELFGLASSIAIKSYFFSNRPIMPILINETNGSKQIPALLVGIIACIFAIFHYOQTANA 458
QY	177 PEKKTIGEVAATGMASHRNPLLAGAAFGSLVPLTFHWFGESVLVSRAVSGHPHGDPNP 236
DB	455 VEHFLRKTKTA-----LSAALSIGTVLFCHHTFLCDSTYLMTSMNGYPRKGPQTP 510
QY	237 FCGAVLLCIASGLMPSCIMEFRGTGLIMWVTSATSAAGLLYT-HTWAAAVSGCFEAIFT- 294
DB	511 HGAVSIVVASICAVLAVPARYOGSAFELLGFVLACFSGSYMVIINHGCSSLGLIFTSYVL 570
QY	295 -----ASMMPQTLGHILNSTGNPKGTMTIMIFYELLEIFFCAMCTAKFYPGGVYAARESD 350
DB	571 ITSEFSI-----RSSYSPPAKWGGAFLVYIILSLAHWWVAALAEFPGGITLERFS 623
QY	351 VLIQTMMILTIQLNMDF-----GPAKNLDLLIQTNSSKVLEFKSEKYMKLEFMWL 400
DB	624 YIL-----IFIGNNLALVAPYAGSESKEPKADSSVDINQDSYSRRSFKNS-----LL 674
QY	401 VGVGILGLGLR 411
DB	675 TGFCIALMALWK 685

## RESULT 6

Q95125 PRELIMINARY: PRT: 279 AA.  
 AC Q95125: 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DE 01-MAY-2000 (Tremblrel. 13, last annotation update)  
 GN SCJ21.12C.  
 OS Streptomyces coelicolor.  
 OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
 OC Actinomycetales: Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 P Seeger K.J., Harris D.;  
 Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Parkhill J., Barrett B.G., Rajandream M.A.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.;  
 RA Kinsahl H., Hopwood D.A., Denapate D., Eichner A., Cullum J.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 3 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL109747; CAB52358.1;  
 SO SEQUENCE 279 AA; 30031 MW; B16719A1698580C0 CRC64;

Query Match 3.88; Score 141.5; DB 2; Length 279;  
 Best Local Similarity 24.98; Pred. No. 0.013;  
 Matches 63; Conservative 36; Mismatches 91; Indels 63; Gaps 12;

QY 447 LERSAHLNENGTADITLESASPKYMGNDLTMW-----LGEKLFYTFDG----- 494  
 DB 58 LRRANVTRKSGADVYGOEVD--KHYASRD--WAQPAELALLGTHYVFGANIDNS 112  
 QY 495 --PSTRYHWGIMALSRPIYKSEH-HLPSPEGEIAPAITLVNIGSKLVDFVYTHF- 549  
 DB 113 PPAQGHVQGTALISRYTPASDNTWLYKSPGQEGRLHATLDVGNKKVEFYNTILA 172  
 QY 550 -GNHEDDLRLQAVSKLSSNOYIFIGYTSAPGSRDYDLDTHGNVKKDIDSTH 608  
 DB 173 AGSQADRLQQAQVVDLIGTRKPG--ILVGFENALPAABE-----SRPLQNAVY 219  
 QY 609 DRWCYIYVRG-----LIRLGVARISHAELSDSEIQMAKFRIPDPPTVYRDNQ 656  
 DB 220 DAMAKSLIARQDGYTPAQSPFERIDLITYAT--RRVTPLYVAQVVK-----DDP----- 265  
 QY 657 KVIDHREVSEKI 669  
 DB 266 -AASDHRELLGKV 277

## RESULT 7

Q95072 PRELIMINARY: PRT: 583 AA.  
 AC Q95072: 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 GN NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3).  
 OS Hyaloraphidium curvatum.  
 OC Eukaryota; Fungi; Fungi Incertae sedis; Hyaloraphidium.  
 NCBI\_TaxID=82268;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Forget L., Ustinova J., Wang Z., Hues V.A.R., Lang F.B.F.;  
 RT "Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing,  
 and an evolutionary link to lower fungi.";  
 RL Mol. Biol. Evol. 0:0-0(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lang F.B.F.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF402142; AAK83434.1;  
 KW Oxidoreductase; Mitochondrion.  
 SO SEQUENCE 583 AA; 64424 MW; 2B56BD4847987022 CRC64;

Query Match 3.58; Score 132; DB 8; Length 583;  
 Best Local Similarity 19.48; Pred. No. 0.18;  
 Matches 116; Conservative 91; Mismatches 194; Indels 196; Gaps 28;

QY 8 ILLESLLGCVSWSLYHDLGPMIYF--PLQTLLELTGLEGSIAFLSPFLTTPFMKLVN 65  
 DB 5 IITPLGLSSSG--LGRWIGYSGGPLTLVLMGV-----TLALVLCGYEIIIF 52  
 QY 66 KKMMLTL--LRITIGSI--ASFQAPNAKLRLMVLALGVSSSLIYQAVTWMSGS-HIQ 118  
 DB 53 HOSSLPLPLPSWTLGLWLDMSFIMDELSIMMIPCTVSLVHMAIGMHGDPHIQ 112  
 QY 119 R--YLRIWGFIL-----GOIVLVLRIMYTSINPIWSYQMSKVI 156  
 DB 113 RFFSYLSLFTFFMLMTYADWMLLFTIGMEVGLVSYLLIGFWTRLR--AGAAALQAF 169  
 QY 157 LTLALATLDRIGTDCGCKPEEKKTGEVATGMSRPWMLAGAFSLVLTWVGEV 216  
 DB 170 L--MNRIGDPTG-----LFLAMA-----IAITMLGDL 193  
 QY 217 SLVSRMAVSGHPHPGDPNPRGAVILCLASGLMPLSCMLFRGTGLMTWVTGTAAGL 276  
 DB 194 EFTFFALLPMPN-----YMTALIGLMLLVTK-SGL 228  
 QY 277 YLHTW-AAVSGVFAIFTASMPOTLGLINSCTNPKTMTIMIFYLE--IF----- 328  
 DB 229 GLHMLPLPAMG-----PTVPSALIHAA--TWVATGIVLLRFNMLFGFSP 322  
 QY 329 FCANCTAFKVPYGVYARSDV-----LLGTMLITIGMLFGPKKNLDLLQ 378  
 DB 273 ILIMGATATLFAVYGVYOTDLKRTIAYSTTSOLGYWVLAGCG----- 317  
 QY 379 KNSKVLFRKSEKMKLFMLLVGVGLIGLRKAYE--RKLGVAPTKEV----- 428  
 DB 318 OYGLALHLVHAFFKALLFLSAG-----SVHAHDEODIRKMGIMSPVYITSM 371  
 QY 429 ---SAIWPFRFGYDNEGM---SLERSAHLNENGTADITLES-----DAS 470  
 DB 372 VGSLSVALPFLTGYISDFILQTAFTGYTLMG-LGAATYATYSLKLHRYFWLPPQS 430  
 QY 471 KPYGNDLIMLGEKLGFTDPSRTRYHWGIMALSRPIYKSEHHLPSPEGEI 527  
 DB 431 KVLGAHPESAMMLIPTTLTFLSIS---WGYLEAQAHAALAPAMVTLFLLPQNOI 482

## RESULT 8

P95356 PRELIMINARY: PRT: 396 AA.  
 AC P95356: 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 GN ORF396 PROTEIN.  
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
 OC Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NCBI\_TaxID=316;  
 RN [1]

	RESULT 9		
RT	079569		
RA	079569	PRELIMINARY:	PRT: 609 AA.
RR	079569:		
DT	01-NOV-1998 (TREMBlrel. 08, Created)		
DT	01-NOV-1998 (TREMBlrel. 08, Last sequence update)		
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)		
DE	NADH DEHYDROGENASE SUBUNIT 5.		
GN	NADH5.		
OS	Musculus manazo.		
OC	Mitochondrion.		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;		
OC	Elaeombranchii; Galeomorphii; Galeoidea; Carchariniiformes; Triakidae		
CC	Musculus.		
OX	NCBI_TaxID=79736;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER.		
RX	MEDLINE=99083431; PubMed=9866199;		
RT	Cao Y., Waddell P.U., Okada N., Hasegawa M.;		
RA	"The complete mitochondrial DNA sequence of the shark (Musculus		

```

0Y 274 GLLY-----LHTW-----AAVSGCVFAI 292
      : : : : : : : : : : : : : : : : : :
Db 173 AVIYRKIGDVGLLSMANLALYNSEWEIHOFLFSKKNDDLPLPLGLVLAAGAAGKAOGL 232
      : : : : : : : : : : : : : : : : : :
0Y 293 F-----TASMMPOQLGHLINSGTNPCKTAMIFYLEIF-----FCAMCTAFK 337
      : : : : : : : : : : : : : : : : : :
Db 233 HPMLSABEGPFPYVALLHS-----STWVAGIFLLIRLHPLOQNKILTFVCLGALT 287
      : : : : : : : : : : : : : : : : : :
0Y 338 FVPGGVYARERDV-----LLGTMMILIGIMLFGRPKNIDLLQTKNSKYLFR 387
      : : : : : : : : : : : : : : : : : :
Db 288 TLETATCALTOIDUKIKVAFTSSQLGIMWTIGLN-----OPOLAFIHI 332
      : : : : : : : : : : : : : : : : : :
0Y 388 KSEKYMKTFLMILYVGLGLGLRKAYE---RKLG---KVAPKVEVSAIWPFRFGYD 440
      : : : : : : : : : : : : : : : : : :
Db 333 CHNAFFKAMFLFCG-----SIHSLEBODIRKMGGLKILPPTTSILIGS----- 380
      : : : : : : : : : : : : : : : : : :
0Y 441 NEGWSLERSAHLNETGADFIT--LLESDAKPYVGNNDLTMWLGCKLGYTDFGPGSTRY 499
      : : : : : : : : : : : : : : : : : :
Db 381 -----LALTGMPLSGCFSSKDALIESNFTSHLNAALILTVATSF---TAI 424
      : : : : : : : : : : : : : : : : : :
0Y 500 HTWGMIN--ALSKYP-----IYKSHHLLRSP-----EGELAPATLTVNINSKGLVYD 546
      : : : : : : : : : : : : : : : : : :
Db 425 YSLRIELFALMAYPRPNTLSPINENNPLVYINIKRLAYGSIIAGLITLNLNT----- 476
      : : : : : : : : : : : : : : : : : :
0Y 547 THFGNHEDDLDRKLOAIYAVSKLLKSSNOVIPLGYTSPAGSRDYLQLOTEHGYNKVIDST 606
      : : : : : : : : : : : : : : : : : :
Db 477 -----PRKTOIMNSPLKLKLSALLVTIMGILLA-----LELTNLTN----- 512
      : : : : : : : : : : : : : : : : : :
0Y 607 DHDRCXEYIMYRGLIR-LGY-ARISHAELSDSEIOMARF---RIPDPPTNRDNOKYVID 661
      : : : : : : : : : : : : : : : : : :
Db 513 SHFKTNPLRLHYHHFNSLGLYFPSIITHRLLPKLSLWMAQYIETHLDQYMNKEIGPKSMIL 572
      : : : : : : : : : : : : : : : : : :
0Y 662 HREVESEKITHNFRFGSYK 679
      : : : : : : : : : : : : : : : : : :
Db 573 QOTSLIKSTOPQOGLIK 590
      : : : : : : : : : : : : : : : : : :

```

RESULT 10	
Q91VY8	
ID Q91VY8	PRELIMINARY;
AC Q91VY8;	PRT; 269 AA.

```

DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AQUAPORIN 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST TUMOR;
RA Strassberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007125; AA07125.1;
KW Porin.
SQ SEQUENCE 269 AA; 28823 MW; 80534DEAB78AB5E7 CRC64;

Query Match 3.3%; Score 123.5; DB 11; Length 269;
Best Local Similarity 22.8%; Pred. No. 0.27;
Matches 60; Conservative 36; Mismatches 82; Indels 85; Gaps 11;

QY 60 FWRKLVNKKWM-LTLRLITIGTSGASFOAPNAKRLMW-----LALGVSSSLIVQAVTW 111
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 10 FWRVAVAEFLAMTLFVITISGALGFNPLERNQTLVDNKKVSLAFGLATLAQSVGH 69
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 112 WSGSHLQRYLRIMGFILGQIVLVVLRIMY-----TSLNP1WSYQMSNK----- 154
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 70 ISGAHLNPATVL-GLLSQCSILRAVYIIAQCGAIVATALLSGITSSLVDSGLGRND 128
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 155 -----VILTSAITLDRIGTDGCSKPEEKKTEGVTGASRRNWL 196
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 129 LANGVNSGGQIGIIEITGLQVLVLTTRRRDLGGSAP-----LALGSLVALGHL 181
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 197 LA-----GAFGSLV-----FLTHWVGEVSLVRMVGSHPHRPPDPNPF-GGAV 241
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 182 LAIDYTGCSINPARSFGSAVITRNFSNHWIT-----WV-----GPFIGAL 222
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 242 LCLASGLMLPSCILMFRGTGLIW 264
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 223 AVLIDYDFILAPRSSDFTDRMKVW 245
   || : || : || : || : || : || : || : || : || : || : || : || : ||

RESULT 11
Q9GJX6 PRELIMINARY; PRT; 636 AA.
ID 09GJX6;
AC 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SODIUM IODIDE SYMPORTER.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RA Selmi-Ruby S.; Rousset B.;
RL Cloning of the pig sodium iodide symporter: Multiple isoforms
   generated by alternative splicing.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ276292; CAC17816.1;
DR InterPro: IPR001734; Na_solut_symport.
DR Pfam: PF00474; SSF: 1.
DR PROSITE: PS00456; NA_SOLUT_SYMP_1; UNKNOWN_1.
DR PROSITE: PS50283; NA_SOLUT_SYMP_3; 1.
SQ SEQUENCE 636 AA; 67165 MW; F80BBFD05152668F CRC64;

Query Match 3.3%; Score 122; DB 6; Length 636;
Best Local Similarity 22.4%; Pred. No. 1.1;
Matches 71; Conservative 43; Mismatches 103; Indels 100; Gaps 16;

QY 15 GCVSWSLYHDLGPMI-----YFPLQTL---ELTGLEGSIAFLSPILFTTPFWK 62
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 300 GIWFALYVDCDPLLAGHISAPDQYMLVDIFEDLPVPG-----LFLACAVSGT 351
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 63 LVNKKMWLTLLRITIGTSGASFOAPN-AKRLMWALGV-----SSSLIVQAV-TWMSG 115
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 352 LSTASTSINMAAVTVDDLKPRPLNAPRRVIISGSLISGACTYAAALSSLLGG 411
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 116 HLORYLRIMGFILGQIVL-----VVLRIWY---TSLNP1WSY 149
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 412 VLQGSFTVMGVISGPIILGAVLVGLFPLSCNTSGVLSGLAAGLALSLVAVAGASLYPPSAQ 471
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 150 QMSNKVLLTSAITATLDRIGTDGCSKPEEKKTG-----EVATGMASRRNWL 198
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 472 SMG---VLPSA-----AGCALPTANASGLQDPVLAVNASSYASSLETDPDQPL 518
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 199 GAAGFSLVFLTHWVGEVSLVRMVGSHPHRPPDPNPFGLAVILCLASGLMLPSCLMFR 258
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 519 AASFVAISLYLYGALGLTSLT-----LCGALISCL-TGPTRSAL--- 557
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 259 GTGLIMW-VTGTASAA 273
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 558 GPGLIMWDLRQTASVA 574
   || : || : || : || : || : || : || : || : || : || : || : || : ||

RESULT 12
Q9GJX7 PRELIMINARY; PRT; 643 AA.
ID 09GJX7;
AC 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SODIUM IODIDE SYMPORTER.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RA Selmi-Ruby S.; Rousset B.;
RL Cloning of porcine sodium iodide symporter.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ276242; CAC17810.1;
DR InterPro: IPR001734; Na_solut_symport.
DR Pfam: PF00474; SSF: 1.
DR PROSITE: PS00456; NA_SOLUT_SYMP_1; UNKNOWN_1.
DR PROSITE: PS50283; NA_SOLUT_SYMP_3; 1.
SQ SEQUENCE 643 AA; 68127 MW; 29B111986C8E8443 CRC64;

Query Match 3.3%; Score 122; DB 6; Length 643;
Best Local Similarity 22.4%; Pred. No. 1.1;
Matches 71; Conservative 43; Mismatches 103; Indels 100; Gaps 16;

QY 15 GCVSWSLYHDLGPMI-----YFPLQTL---ELTGLEGSIAFLSPILFTTPFWK 62
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 300 GIWFALYVDCDPLLAGHISAPDQYMLVDIFEDLPVPG-----LFLACAVSGT 351
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 63 LVNKKMWLTLLRITIGTSGASFOAPN-AKRLMWALGV-----SSSLIVQAV-TWMSG 115
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 352 LSTASTSINMAAVTVDDLKPRPLNAPRRVIISGSLISGACTYAAALSSLLGG 411
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 116 HLORYLRIMGFILGQIVL-----VVLRIWY---TSLNP1WSY 149
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 412 VLQGSFTVMGVISGPIILGAVLVGLFPLSCNTSGVLSGLAAGLALSLVAVAGASLYPPSAQ 471
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 150 QMSNKVLLTSAITATLDRIGTDGCSKPEEKKTG-----EVATGMASRRNWL 198
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 472 SMG---VLPSA-----AGCALPTANASGLQDPVLAVNASSYASSLETDPDQPL 518
   || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 199 GAAGFSLVFLTHWVGEVSLVRMVGSHPHRPPDPNPFGLAVILCLASGLMLPSCLMFR 258
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 519 AASFVAISLYLYGALGLTSLT-----LCGALISCL-TGPTRSAL--- 557
   || : || : || : || : || : || : || : || : || : || : || : || : ||

```

Wed Oct 23 09:09:29 2002

QY 259 GTGILMW-VTGTASAA 273  
 Db 558 GPGILMWDLTRGTASVA 574

## RESULT 13

Q9UYPS PRELIMINARY; PRT: 766 AA.  
 ID Q9UYPS  
 AC Q9UYPS:  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE TRANSMEMBRANE OLIGOSACCHARYL TRANSFERASE, PUTATIVE.  
 GN PAB0974.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN NCBI\_Taxid=29292;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
 DR EMBL: AJ248287; CAB50367.1; -  
 DR InterPro: IPR003674; OTase\_STT3.  
 DR Pfam: PF02516; STT3; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 766 AA; 87894 MW; C7A7E52701819CE4 CRC64;

Query Match 3.3%; Score 122; DB 17; Length 766;  
 Best Local Similarity 23.5%; Pred. No. 1.4; Mismatches 153; Indels 90; Gaps 19;

QY 286 SGCVALEFASMPQTLGLHNSGTNGKMTIAMIFYLEIFCAGCAFEKVGVA 345  
 Db 51 AGWNEFFETIANGP--WGFOIKSFHPLGLMTPAYIRFLKVEGISVQTEFKTP----- 103  
 QY 346 REKSDVLTGTMMLI---IGLNLFGPKKNDLLQTNSSVLEPKSEKYM---LEFLW 398  
 Db 104 -----VFGLTTFVFYISLTKLGGKAPFASFLALSFGHIFRSMANYRGDNMLFW 158  
 QY 399 LLVGVGLGLGRHAKAYERKIGVAPT--KEVSAIMP-----FRFGYDNE-----GMS 446  
 Db 159 YSLALAIATVLRKGRVLYVPLTLAGSISVFQWATYPLFVLSLNCVFLAIGSFL 218  
 QY 447 LERSAHLNETGADFTILSDASKPYMGNNDLTMMLGEKIGFYTDGSPSTRYHWGIMA 506  
 Db 219 LDKRKNFLD---SFIIIL-STAFGAIIAN---YLGEKFGY-----GMLG 255  
 QY 507 LSRPIYKSEHHLPSPEGELAPITLVNISCKLVD---FVYVHFGNHEDDLDRKIQAI 563  
 Db 256 YNRHIVYSK-----LGIKL-GKIRDAVLEFHLHY-LVPISLGLLITLL 296  
 QY 564 AVSLKSSNQ--VIFLGYT-----SAPGSRDYQLTEHGNVNDIDSTD---HDR 610  
 Db 297 FLSRFVSKRAKAGVIGLGVSTLILLLKFPALRGGLGIFDMKSPIMETRTNHD 356  
 QY 611 WCEYIMYRGLRIGYARISHAELSDSE 637  
 Db 357 WKAFSISIFLLPLFLRF-HPEKVKTE 382

RESULT 14  
 Q36149 PRELIMINARY; PRT: 591 AA.  
 ID Q36149  
 AC Q36149:  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT 5.  
 GN ND5.  
 OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).

OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.  
 RN NCBI\_Taxid=34903;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART;  
 RX MEDLINE=97054595; PubMed=8898892;  
 RA Cai Q., Storey K.B.;  
 RT "Anoxia-induced gene expression in turtle heart. Upregulation of  
 RT mitochondrial genes for NADH-ubiquinone oxidoreductase subunit 5 and  
 RT cytochrome c oxidase subunit 1."  
 RL Eur. J. Biochem. 241:83-92(1996).  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 DR EMBL: U49048; AAC60055.1; -  
 DR InterPro: IPR003916; NADHdh\_oxidctse5.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR InterPro: IPR001412; tRNA-synt\_1.  
 DR Pfam: PF00361; oxidored\_q1; 1.  
 DR Pfam: PF00662; oxidored\_q1\_N; 1.  
 DR PRINTS: PR01434; NADHGNASE5.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_1; UNKNOWN.1.  
 DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
 SQ SEQUENCE 591 AA; 66062 MW; 1C577C2B77FE901A CRC64;

Query Match 3.3%; Score 121.5; DB 8; Length 591;  
 Best Local Similarity 20.0%; Pred. No. 1.1; Mismatches 145; Conservative 86; Mismatches 226; Indels 269; Gaps 36;

QY 1 MTSIMREILLESLLGCWSLYHDGPMY--VEPLQTL---ELTGLEGFS----- 47  
 Db 5 MSLSKALLIKKREISTYKMTFFISLFLHSMLEPEQTLCDVILGLSLSIKINFHDQ 64  
 QY 48 -AELSPFTLTPPEWKLNV-KKMLTLRLITIGSIASFQAPNAK-----LRMLVLA 97  
 Db 65 YSAIFEMIALYVT--WNILQYSRWY-----MDNPHIDKFFKYLFLFMMAM 109  
 QY 98 LGVSSSLIQAQVWWSHLORYIRINGFILQIVLVVLRITWSLNPISYOMSNVIL 157  
 Db 110 ILITSNLFFQFIFIGWES-----SNHILL 133  
 QY 158 T-----LSAIVTL--DRIGTDGCKSPREKKTGEVATGMSRPMLLAGAFGS 204  
 Db 134 THMNVTRREDTLASALAIITNRIGDLG-----LYIGMV--WLSKT--NS 175  
 QY 205 LVFLTHWVGEVSLVSRWAVSGHPDPNPFEGAVLLCLASGLMLPSCILWFGTLIW 264  
 Db 176 LNF-----QPSYTPN--SLVMLIA-----LIL 196  
 QY 265 WVTGTASAGLLYHTW-AAAVSCVFAITFASMPQTLGLHNSGTNPGKTIAMIFY 323  
 Db 197 AATGKSAQFG--LHPMLPAMEG-----PPVVALHS-----STNVVAGIFL 237  
 QY 324 LLEIF-----FCAMCTAFKFPVGGVYARESDV-----LGTMMILIGL 362  
 Db 238 LIRHPLASNTPLSTCLGATITTPFASLSAQNDIKKIIVYSTQGLMAMVITGL 297  
 QY 363 NMLFGPKKNDLLQTNSSKVLFRKSEKYMKLFLMLLVGVGLGLGRHAYERKIGKY 422  
 Db 298 N-----QPDIAF-----LHICMNAFFKAMLPFACAGRIHSLG--NQDIRKMG 340  
 QY 423 -APTKVSAI-----WPRFGYDNEGWSLSERSAHLNETGADFTILSDASKP 472  
 Db 341 HKPLPITSACLTIGNMALMGMPFLTGYSKAIIETSTYLSNC--ALLTLMATSTFVAV 399  
 QY 473 YMGNNDLTMMLGEKL-----GFYTDGSPSTRYHWGIMALSRYPKSEHHLPS 523  
 Db 400 YTLAIMLVQGTGHLOQSTYILHNIFTIAMPITRLATGSIAMA--GMLITKINQMNT 456  
 QY 524 EGETAPAI--TLTVNISGKLVDVVTVHFGNHEDDLDRKIQAIIVSKLK---SSSNV 576  
 Db 457 TMTPMWIKIKIALTVTVIGLLVGL-----KVCVMAVTKPKPKYKTCDDPTSM 502

```

QY 577 IFGLYTSAPC-SRDYQLTTEHGNVKIDSYDHRMCEY-----IMY 617
DB 503 AGLOLSPSPPOYIKKYLKMAKXIKTHLADTS-----WPEYLCPOXYTAKMKTPTMFTSLMO 558
QY 618 RGLIRL 623
DB 559 KGLIKI 564

RESULT 15
Q9B6Y4 PRELIMINARY; PRT; 605 AA.
AC Q9B6Y4;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, last sequence update)
DT 01-DEC-2001 (TREMblrel, 19, last annotation update)
NM NADH DEHYDROGENASE SUBUNIT 5.
OG Casuarinus benneitii (Dwarf cassowary).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauilia; Aves; Palaeognathae; Casuariiformes; Casuariidae;
OC Casuarinus.
OX NCBI_TaxID=30463;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RX MEDLINE=21085666; PubMed=11217857;
RA Cooper A., Lalueza-Fox C., Anderson S., Rambaut A., Austin J.,
RA Ward R.;
RT "Complete mitochondrial genome sequences of two extinct moas clarify
RT ratite evolution.";
RL Nature 409:704-707(2001).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
DR EMBL: AY016011; AAK08552.1;
DR InterPro: IPR003916; NADhub_oxidctcs5.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; Oxidored_q1_N.
DR Pfam: PF00662; Oxidored_q1_N: 1.
DR PRINTS: PR01434; NADHDHGNASE5.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
KW SEQUENCE 605 AA; 66408 MW; 41B286D9EC828A26 CRC64;

Query Match 3.3%; Score 121; DB 8; Length 605;
St Local Similarity 19.6%; Pred. No. 1.2;
Matches 139; Conservative 87; Mismatches 230; Indels 254; Gaps 34;

QY 27 PMIYFPLQLELTGLGEGFSI-----AFLSPILFTTPMKLVNKKMMLTLRLRIITIG 79
DB 35 PMSTSTVKTAFPSLIPMSIFISGFETIVSOVMYWNFTPNFKIP----- 79
QY 80 SIASFQAPNALKRLMVLALGVSSSLIYQAVYWMSSG--HLQR-YLRIMGFTLGQIVLV- 135
DB 80 -LSFKMDQYSMLFFPALFVYWS--ILOFAVYMASEPHTKFFIYLTFLIAMLTFLIA 136
QY 136 -----LRIWYSLNPI-----WSYQMSNKYILTLISAATLDRIGTDGCSKPEEKTGE 184
DB 137 NNMFMFLPGMGVCGIMSTFLIGMWHGRAENTALQAV-IYNRIGDYG----- 183
QY 185 VATGMASRPNNLAGAAGSLVFLTHWVGEVSLVSRMAVSGHPHGPDPNPFCAVLLC 244
DB 184 -----LILSNMWL-----ASTINWEIQASVETQMP----- 210
QY 245 IASGLMPLSCIMFRGTGLIMVNTGTASAGLLYHTW-AAVSGCVRAIFTASMMPTLG 303
DB 211 -----MLPL-----LGLILAATGSAQFG--LHPWLPAMEG-----PTPVVS 245
QY 304 HLINSGTNPCKTMTIAMIFYLLE-----IFPCAMCTAFKFPVPGVYARERSDV- 351
DB 246 ALLHS-----STMYVAGIFLLIRTHPLLTNNOTALTTCCLGALSTLFAATCALQNDIK 300
QY 352 -----LLGTNMLIIGLMLFGPKKNDLLLTQTKNSSKVLFRKSEKYMKFLMLLVG 402

```

```

DB 301 KIIAFTSSQIGLMMVTIGLNL-----POLAFLHISYHAFKAMLFICSG 345
QY 403 VGLLGLGRHKAYERKLG---KVAPTKEVSAI-----WPFRRFGDNGW---SLERS 450
DB 346 SIHSLGGEODI--RKMGLQKLIPTTACLTIGSLAMKTPPLAGYSKDLIESLNS 403
QY 451 AHLNETGADFTILLESDAKSPKMGNNDLTPWMLGEKGFYTDGSPSTRYHTWGMALRSY 510
DB 404 --YLN-TWALLTLTLATSFATY-----SIRWTLAVQAG-----TRI 438
QY 511 PIKSEHLLPSPGELAPAT-----LTVNISGLV 542
DB 439 PPI-----TPINENSPAVTNPITRLAVGSIAGLLITSNVLPRTPTMPTMVKMA 490
QY 543 DFVYTHRGNHEDDLRKLQIAVSKLKS-----SNQVFLGYITSAPGSROYLOLTH 597
DB 491 AIVYTLG-----IMLDELNLTHNLTPSKONTYSNFSITLGFNPLHRLNPMMLLN 545
QY 598 GNVKDIDSTDRMCEYIMYKGLIRLG-YARISHAELSDSEIQMAKFRIP 646
DB 546 GQ-----KIAH-----LIDLSMYKKMGPEGLADQLMAKKTSTP 580

```

Search completed: October 21, 2002, 22:10:57  
 Job time : 68 secs